Essays in the Economics of Procurement
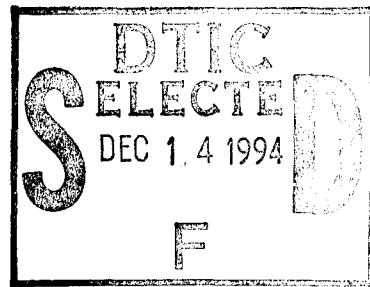
Edited by Anthony G. Bower and James N. Dertouzos

19941207 008

DTIC
SELECTED
DEC 1 4 1994
F

RAND

National
Defense
Research
Institute

This volume of essays is part of a larger project on the economics of defense acquisition. This work was sponsored by the Director, Program Analysis and Evaluation, and was carried out in the Acquisition and Support Policy Program of RAND's National Defense Research Institute, a federally funded research and development center supported by the Secretary of Defense, the Joint Staff, and the defense agencies.

The separate essays in this volume summarize earlier RAND reports and reprint articles that have appeared in the professional literature. The following publications, among others, were fully or partially supported by this project:

James J. Anton and Dennis A. Yao, "Measuring the Effectiveness of Competition in Defense Procurement: A Survey of the Empirical Literature," *Journal of Policy Analysis and Management*, 1990.

Anthony G. Bower, "Procurement Policy and Contracting Efficiency," *International Economic Review*, 1993.

Anthony G. Bower and Kent Osband, "When More Is Less: Defense Profit Policy in a Competitive Environment," *The RAND Journal of Economics*, 1991.

Frank Camm, *How DoD Policy Affects Private Expenditure on Independent Research and Development: A Comparison of Empirical Studies*, RAND, N-2834-OSD, 1989.

Yeon-Koo Che, "Design Competition Through Multi-Dimensional Auctions," *The RAND Journal of Economics*, 1993.

Stefan Reichelstein, "Constructing Incentive Schemes for Government Contracts: An Application of Agency Theory," *The Accounting Review*, October 1992.

William P. Rogerson, *An Economic Framework for Analyzing DoD Profit Policy*, RAND, R-3860-PA&E, 1992.

William P. Rogerson, "Incentives, the Budgetary Process, and Inefficiently Low Production Rates in Defense Procurement," *Defence Economics*, 1993.

William P. Rogerson, *Profit Regulation of Defense Contractors and Prizes for Innovation*, RAND, R-3635-PA&E, 1992. A shorter version also appeared in the *Journal of Political Economy*, Vol. 97, 1989.

William P. Rogerson, "Quality vs. Quantity in Military Procurement," *American Economic Review*, Vol. 80, 1990.

Rachel Schmidt, *Defense Profit Policy and Capital Investment*, RAND Graduate School Dissertation, 1993.

Lars Stole, "Information Expropriation and Moral Hazard in Optimal Second Source Auctions," *Journal of Public Economics*, forthcoming.

CONTENTS

FIGURES

TABLES

ACKNOWLEDGMENTS

INTRODUCTION
by James N. Dertouzos

At the peak of defense spending during the mid 1980s, over 60 percent of the Department of Defense's (DoD's) annual outlays of nearly $300 billion was spent on research, development, testing, and production contracts with private sector firms.[1] By necessity, the award, administration, and completion of these contracts is subjected to an extensive, complex, and idiosyncratic series of rules and regulations. This process, which governs the interaction between DoD and contractors, appears to be fertile ground for economic analysis.

However, insufficient scholarly attention has been directed toward the analysis of these heavily regulated markets. Although significant portions of the applied economics literature address issues related to public utilities, health care, and telecommunications, many fewer contributions to defense procurement have emerged. This dearth of relevant research may not be surprising given that procurement regulation is extremely complex, diffuse, and decentralized. In addition, products and services purchased are highly differentiated and multidimensional, making generalizations quite problematic. As a result, analysts are faced with significant obstacles to acquiring reliable data and integrating information on the regulatory process, institutions, and structure of the defense industry.

This is unfortunate, because recent advances in the literature, including general theories of incentive contracting and more specific applications of microeconomic concepts to procurement issues, have potentially important implications for the regulation of the defense industry. However, previous research on the economics of procurement has not, in general, been used in the design of specific government policies. This stems from a variety of reasons. To begin with, much of this work is published in scholarly journals and remains virtually inaccessible to most practitioners. These articles are prepared for academic audiences that are more interested in the technical contributions of the research than in policy relevance. Even if a dedicated practitioner wished to survey this literature to gain theoretical insights into actual problems, the models rarely incorporate real-world attributes that characterize defense procurement. This is not surprising, since scholars do not

---

[1]Even with subsequent budget declines amounting to over 20 percent in nominal terms, the magnitude of taxpayer resources allocated via the procurement process remains very large.

have the opportunity or incentives to invest in learning the specifics about the relevant institutions and regulations.

The contributions compiled in this volume represent an initial effort at narrowing the rather wide gap between economic theory and procurement practice. The included papers and summarized reports provide important insights as well as a methodological foundation that can be further adapted to the peculiar institutional, political, and technological circumstances characterizing defense production. Ultimately, analyses built on this foundation should improve the process by which the government purchases goods and services from the private sector.

We will begin this introduction with a brief overview of the primary policy tools available to the government in affecting procurement outcomes. This discussion is meant to set the stage for a summary of the essays contained in this volume and their insights concerning the appropriate design of these policies. Next, we will identify a subset of the fundamental elements of procurement, emphasizing some key factors that should be considered if economic analysis is to play a useful role in the design of government policies. We will then discuss the papers comprised by this volume, describe the essential modelling assumptions made in each, summarize the key findings, and draw preliminary inferences for procurement policy.

1. THE REGULATORY FRAMEWORK

Procurement policy has numerous and, sometimes, conflicting objectives, including the efficient production of low-cost, high-quality weapons systems and the provision of a "fair" rate of profit for contractors. Other goals might include maintaining a strong industrial base, high levels of employment, and clear technological advantages vis-à-vis international competitors. Given these multiple goals of procurement policy and the economic barriers to achieving them, what are some of the key policy levers available to decisionmakers?

Promoting Competition

Given that only a handful of firms can produce major defense systems and that scale economies limit the number of market participants, a variety of government options have been developed to promote competition. It is important to distinguish between policies designed to inspire competition for the market (i.e., for the initial contract) and those meant to stimulate *ex post* competition such as dual sourcing, teaming, and other forms of mandated technology transfer. Rigorous, ongoing competition once a program has been established has often been impractical because of purported scale economies, sunk costs distinguishing incumbents from potential rivals, and technological advantages gained from experience (i.e., "learning by doing"). Regardless of whether procompetitive policies are pursued or not, it is essential that government policymakers remain aware of the fact that the terms of a contract as well as the source selection procedure for that contract will affect anticipatory behavior of firms (e.g., independent research and development, physical and human capi-

tal investments, and the degree of learning-by-doing) on current as well as future procurements.

Profit Policy

When contractors are not selected through price competition, extensive direct control over the firm is exercised. It is less well known that, even in cases where a firm-fixed-price contract has been awarded via competitive bidding, government auditors will limit *ex post* profits to a "fair" rate of return. It is interesting to note that the vast majority of large contracts are ultimately negotiated according to the "weighted guidelines," published by the DoD as a guide to procurement officers. In such instances, profit is calculated as a complex function of various operating and capital cost components. Allowable profit depends on a number of factors, including contractor performance (is the project completed on schedule and according to specifications?), input mix (until recently, working capital was treated much more favorably than facilities capital, for example), and perceived levels of technical and contract risk (fixed-price contracts are inherently more risky).

A major component of allowable expenses consists of overhead costs, which are usually allocated to specific contracts on the basis of fairly simplistic accounting rules that give contractors significant discretion across product lines. It will be important to note that what policymakers often interpret as the price or economic cost implications of alternative procurement strategies may, in fact, merely represent an "accounting shuffle" of overhead costs from one contract to another.

Contract Design

The government employs a variety of contract types. The general forms of these contract alternatives, fixed-price, cost-plus, and incentive contracts, have been analyzed extensively in the literature, at least from a theoretical perspective. In practice, there are strong correlations between the type of contract and project attributes, such as the nature of the competition, degree of technological sophistication, and stage of development, that are related to uncertainty in cost projections. It is also noteworthy that the frequency of contract types that place much of the burden of risk on contractors has increased significantly over the years.

In addition to considering risk, the literature on optimal incentive contracts and auction design emphasizes the tradeoffs stemming from moral hazard and adverse selection.[2] Of course, it is important to recognize that contract forms actually used by

---

[2]Moral hazard refers to a situation in which contractor incentives and government objectives are not well aligned and the government cannot perfectly observe the actions of the contractor. For example, if the firm makes more money by being inefficient and the government cannot monitor effort, then cost-plus contracts may induce suboptimal levels of effort to become productive. Adverse selection arises when the government cannot perfectly observe what "type" of firm it is dealing with. For example, DoD does not know how much a contractor's product costs even though the contractor does. Both terms originated in the insurance context, where their names have more intuitive appeal. For the rest of this volume, just keep in mind that "moral hazard" refers to hidden action and "adverse selection" refers to hidden type (usually cost of the firm).

DoD could well deviate from optimality, narrowly defined, for valid reasons. For example, decisionmakers may wish to promote growth in the defense industrial base or, perhaps, have strong incentives to minimize the burdens of administering complex contracts.

2. OVERVIEW OF DEFENSE PROCUREMENT

Defense procurement has several key elements that should be considered in economic analyses undertaken for the purpose of influencing public policy.

Multiple Regulatory Objectives

It is worth noting that there are multiple goals of procurement that may be inconsistent with one another. The most obvious goal is to facilitate the timely procurement of high-quality weapon systems at minimum cost to the taxpayers. Of course, the quality/cost tradeoffs may be viewed quite differently by different participants in the process (more on this below). In addition to the goals of allocative and technical efficiency, the regulations are designed to provide a "fair" rate of return to contractors. From the politicized perspective of government agents, regulators may be more concerned with apparent returns rather than with long-run economic profits. This distinction may explain why current regulations are applied on an individual contract rather than on a firm level. Next, policymakers are concerned with maintaining a healthy defense industrial base. Especially during the current drawdown, uncertain future national security requirements require sufficient surge capacity. Finally, because appropriations are not fungible across budget categories, administrative expenses and costs of implementing procurement strategies should be considered independently.

The Multiple Players in Defense Procurement

The relative importance of the competing objectives of procurement policy is quite different across the various political and economic participants. The existing regulatory process is diffuse and decisions are often made sequentially by independent agencies having contrasting objectives. For example, the service personnel making design decisions may be less concerned with life-cycle costs related to personnel training or system maintenance. The DoD might well seek to maximize national security goals, while the Congress likely has a broader perspective on national priorities, including domestic as well as defense objectives. The important point is that the hypothesized tradeoffs between aggregate costs and benefits from society's point of view are seldom considered by those decisionmakers influencing outcomes.

Government Contracts Are Interrelated

Individual contracts should not be viewed as separate and independent entities for modelling or policy purposes. To begin with, the development of a complex, high-technology system goes through several stages over a period of decades. These

stages are interdependent and government policies that alter incentives in one stage are likely to influence outcomes in earlier or subsequent periods as well. For example, attractive rates of return in the production phase of a procurement are thought to induce socially desirable rent seeking in the form of research and development. In addition, contractors are simultaneously engaged in multiple projects, both government funded and commercial, at any given point in time. These projects are interlinked via the production process (i.e., economies of scope) as well as because of accounting rules that dictate the allocation of joint costs between economic activities. Finally, historic outcomes will inevitably influence expectations about future outcomes.

Incomplete Information and Uncertainty

The acquisition process is characterized by incomplete information and uncertainty. The government typically has incomplete information about the contractor's actions or type, and procurement policies should recognize this. The theoretical economics literature has made substantial progress in the last decade in characterizing optimal contracts under incomplete information and one major contribution of this volume is to take those insights and apply them to the defense procurement context.

Uncertainty comes in two varieties. The first kind of uncertainty is that encountered in well-understood, but random processes, such as horse races. We do not know who will win the race, but the potential outcomes and odds of each horse can be accurately described. Similarly, some defense procurement situations may be characterized by this type of uncertainty or risk. Procurement policies should make some attempt to mitigate the welfare loss associated with risk (for contractors, government agents, taxpayers, and their representatives). And, as is well discussed in the contracting literature, this additional policy objective creates a dilemma for policymakers wishing to elicit desirable behavior (i.e., reduction of moral hazard) on the part of contractors, award the contract to the most efficient firm (i.e., limit adverse selection), and procure the system at a low (but fair) price.

The second variety of uncertainty is at a more general level. Specifically, high levels of uncertainty about the very rules of the procurement relationship and difficulty in even describing the potential situations that might face DoD and the contractor over long periods of time create problems in even constructing formal models to study procurement. This should not be considered a defect of the formal procurement literature to be presented. If any economic situation is extremely uncertain (as, at times, defense procurement seems to be) then formal models will not be as helpful as less-structured descriptions of potential problems and solutions. However, as always, if the economic situation can be characterized in a reasonably certain way, then formal models can illuminate previously unrecognized tradeoffs and demonstrate the optimality of a particular course of action.

In the above discussion, we have briefly characterized both the major policy levers (inducing competition, promulgating profit policy, and designing contracts) and the defense procurement environment (multiple regulatory objectives, multiple players, interrelated contracts, and incomplete information and uncertainty). The articles in

this volume, discussed in the next section, contribute to our understanding of the relationship between the defense procurement environment and the usefulness of the various policy levers available to the government.

3. OVERVIEW OF RECENT CONTRIBUTIONS TO THE ECONOMICS OF PROCUREMENT

The first contribution to this volume summarizes recent analysis of DoD policies, known as the "weighted guidelines," for negotiating fees (and, therefore, nominal profits) on sole-source contracts for the provision of goods and services.[3] As Nachbar notes, despite a policy emphasis on competitive procurements following the adoption of the 1986 Competition in Contracting Act, over 60 percent of all contract dollars are allocated according to these guidelines.[4] This high percentage stems from the fact that even competitive contracts are often renegotiated following inevitable design changes. In addition, about half of all competitive contracts rely on the weighted guidelines for setting targets or ceilings on allowable profits.

The guidelines specify allowable ranges of markups over costs for a number of factors, including payments for the use of working capital, land, buildings, and equipment. Although determining the "right" level of compensation overall can be controversial (we argue below that it may, in fact, be irrelevant), the basic structure of these payments seems sound. One exception to this conclusion stems from the treatment of certain true economic costs that are not recognized by the guidelines and are, in practice, indistinguishable from pure profits. For example, the guidelines make no adjustment for the time value of money. That is, nominal payments are viewed as being identical regardless of whether they are delivered in the first year or at the end of a multiyear contract.

Perhaps more important, the guidelines provide separate payments for both technical and contractual risk, but the additive structure of these markups is incorrect. That is, the technical risk of experiencing cost overruns vanishes under a cost-plus contract. Such risk becomes real if and only if the firm has to pay for all (in a fixed-price contract) or some (in a cost-sharing contract) of the cost overruns. On the other hand, in the absence of technical risk (such as in the routine manufacturing of simple goods with stable technologies and input prices), the form of the contract does not affect the overall level of risk, which is low in any case. Formally, therefore, the level of overall risk should be expressed as the product of technical and contract risk.

As Nachbar notes, however, the empirical significance of these "flaws" is uncertain. To begin with, the guidelines provide DoD negotiators considerable latitude in choosing markups within rather broad ranges. In theory, government personnel can adjust individual markups within categories to promote virtually any objective, both

---

[3]This contribution by John Nachbar summarizes several related papers, including unpublished mimeos by Kent Osband and a report by William Rogerson (1992a).

[4]In 1983, 75 percent of awards were allocated under negotiated contracts using the weighted guidelines. See Schmidt (1993).

in terms of the desired level of overall compensation or the relative returns to alternative categories of investment. Indeed, there exists no convincing empirical evidence that the structure of the guidelines has significant effects on procurement outcomes.[5] Still, a careful restructuring of profit policy and explicitly allowing for the correct compensation for risk and opportunity cost could, at the very least, reduce the administrative costs of negotiating a contract.

In addition, theoretical work by Rogerson suggests that the treatment of overhead costs by the weighted guidelines may create a variety of distortions that can affect procurement outcomes.[6] This is because a large fraction of costs are difficult to attribute to particular projects and are, instead, grouped in overhead pools. Such indirect costs are generally allocated across products in proportion to direct labor use. To the extent that contractors have discretion in the mix of inputs utilized on a particular project, incentives will exist to alter that mix to increase the allocation of overhead to projects in which price is more responsive to accounting costs. For example, firms will wish to allocate overhead to sole-source, cost-plus contracts and away from commercial or competitive fixed-price contracts. As Rogerson notes, if this incentive is strong enough, a firm may respond by engaging in pure waste, i.e., hiring surplus labor on certain projects, thereby shifting a greater proportion of indirect costs to "price sensitive" contracts. In addition, a firm can alter its labor/capital mix by varying production processes. For example, the firm can automate production on commercial enterprises and on competitive procurements, thereby lowering direct labor on these contracts. In contrast, by reducing its level of subcontracting on cost-plus contracts and bringing more production in house, the firm reduces the use of material and increasing labor in these operations. As a result, the firm is able to allocate overhead (as a percentage of direct labor) to those contracts on which it is possible to recapture such costs.[7]

Thus, the structure of the weighted guidelines can distort firms' choices away from production efficiency by allowing cross-program shifts in accounting costs. In particular, the guidelines could affect subcontracting, automation, and the use of labor. In addition, such shifts distort the relationship between accounting profitability on specific contracts and the true economic profitability of the firm overall. Such distortions may well be serious, and profit policy revisions that serve to reduce the percentage of indirect costs along with the adoption of accounting conventions that reduce flexibility in allocating such costs would be desirable.

Regardless of the structure of the markup policies used to negotiate the fee on a particular contract, it is clear that, when viewed in a dynamic or multiperiod context, this payment may have little to do with classic notions of economic profitability. For example, Rogerson, in the second paper in this volume, argues that contract-specific

---

[5]For example, there is no evidence that rather dramatic changes in the size of the allowable markups to facilities capital have any effect on investment behavior of contractors. See Schmidt (1993).

[6]For details, see Rogerson (1992c).

[7]Rogerson provides a rough estimate of the magnitude of this incentive effect by using data on the overhead pools of four large aerospace firms. He calculates that incurring an extra dollar of direct labor cost on a sole-source procurement can generate up to $1.44 in extra revenue.

profits are best viewed as a reward or prize for innovative activity. Such R&D investments are undertaken in the hopes of winning a profitable production contract. In equilibrium, firms will continue to spend their own money up to the point of expecting a reasonable rate of return on their R&D investment. With competition between firms in the development of innovative product ideas and new technologies, firms can expect no more than a fair return to their efforts over the entire product life cycle. In other words, when one considers the successful as well as the unsuccessful R&D efforts, overall firm profits, at least on average, cannot be excessive in the long run, regardless of the level of the fees offered on a specific contract. As markups rise, competition for the right to earn such higher fees becomes more fierce, and firms spend more money on innovative activity. Taking a broad perspective, profit policy will therefore affect levels of R&D expenditures but have no effects on long-term profits.[8]

In an empirical analysis of stock price fluctuations of competing defense firms subsequent to a contract award, Rogerson is able to obtain an estimate of the real economic profit associated with a particular contract. He finds that implied profits amount to just under 5 percent of total contract value. This implies a modest, but significant economic profit. This result is suggestive but falls short of validating his notion that contract profits are merely a fair rate of return to earlier innovative activity. In future investigations, it would be desirable to link information on earlier R&D expenditures with systematic differences in the size of the implied profit level (as indicated by increases in the stock price of the winning firm as well as symmetric declines in the securities of losing competitors).

The contribution by Bower and Osband again demonstrates the importance of taking a multiperiod perspective on the defense procurement process. In particular, they explicitly recognize the interdependence of serial contracts. For example, even when contracts are awarded through a competitive process, such contracts are often renegotiated during a subsequent phase. In addition, the initial winning firm often builds such an insurmountable technology and cost advantage that competition for the follow-on production work is not feasible. As a result, follow-on contracts are often negotiated with a single source using the weighted guidelines.

Firms are surely aware that winning the initial competition provides an inside track to subsequent contracts in which profits are typically negotiated as a markup over cost. As a result, firm bidding behavior during the initial stages will inevitably consider subsequent profit levels. In fact, less-efficient firms stand to gain more profit in the cost-plus environment and, as a result, are willing to bid quite aggressively during the competitive stages.

The following simple example, outlined in Table 1.1, illustrates the main insight from this characterization. Imagine that there are two firms. The first, the low-cost firm, is

---

[8]Of course, it is possible that high individual profit levels could induce too much innovative behavior, especially if some of it is redundant activity from a social perspective. In addition, some activities that may enhance the probability of winning a profitable contract may not be in society's best interests. For example, excess "gold plating" or intense lobbying may not improve acquisition outcomes relative to the costs of such efforts.

capable of producing a good for $100 in each of two periods. In contrast, the high-cost firm has double the costs, $200 in each period. If the allowable markup is 10 percent, the prospective profit in period 2 will be $10 for the low-cost firm and $20 for the less-efficient firm. This means that the less-efficient firm is willing to bid $180 for the opportunity of making $20 during the negotiated phase. This is because $180 minus the cost of $200 creates a loss of $20 in the first stage, exactly counterbalancing second-period profits. As a result, the more-efficient firm need only go as low as $180 in bidding against its higher-cost rival. For this firm, the competitive bid of $180 yields a first-period profit of $80. Note that the firm earns a total of $90 for both periods. This excess of profits over actual costs is termed the "information rent," stemming from the government's inability to audit costs accurately during the first phase. The size of this rent is related to the wedge between the winner's costs and those of the second-most-efficient firm. It is important to note that the government pays a total of $290, representing contractor profits plus actual costs.

Now, allow for an increase in the allowable markup for the negotiation stage. This means that the less-efficient firm can earn $40 in the second stage, implying that it would bid as low as $160 in the first stage. It is willing to lose more in the first stage to gain access to the higher profits made available during the cost-plus phase. This lower bid in the first stage means that the low-cost firm, to win the contract, has to lower its bid to $160. As a result, its profit falls to $80, even though the allowable markup has increased in the second stage. Interestingly, the government pays a total of $280, representing a reduction in the information rent as well as in the total price.

This inverse relationship between markups and costs does not prevail for all levels of markups. In particular, if the profit advantage enjoyed by the high-cost firm exceeds the cost advantage of the efficient firm in the first stage, then the inefficient firm will win the competition. If the objective is to minimize government costs, then the markup should be increased just to the point where the advantage enjoyed by the low-cost firm is dissipated and information rents eliminated.[9]

Table 1.1

Relationship Between Procurement Costs and Markup

| Type of Firm | Markup = 10% | | Markup = 20% | |
	Period 1 Profits Bid minus Cost	Period 2 Profits Markup x Costs	Period 1 Profits	Period 2 Profits
Low-cost firm, $100 per period	Bid – $100, or 180 – 100 = $80	.1 x $100 = $10	Bid – $100, or 160 – 100 = $60	.2 x $100 = $20
High-cost firm, $200 per period	Bid – $200, or 180 – 200 = –$20	.1 x $200 = $20	Bid – $200, or 160 – 200 = –$40	.2 x $200 = $40

[9]As Bower and Osband demonstrate, the optimal markup will be related to the ratio between total costs incurred in the first versus the second period. For example, if first-period costs represent 10 percent of total costs, then a markup of over 10 percent will reward inefficient firms enough to be able to outbid the low-cost firm, to the detriment of the government decisionmakers.

Bower and Osband generalize their analysis to include incentive effects on learning-by-doing created by the type of contract employed. They note that a higher markup implies a higher level of cost reimbursement and will dull incentives to reduce costs during the life of the contract. The optimal markup strikes the right balance between inducing vigorous bidding competition identified above, and preserving incentives for firms to engage in (costly) learning to lower costs.

This inverse relationship between the markup and price (and profits) occurs because the second-period markup serves as a bidding subsidy that favors a high-cost firm and makes the other firm bid more aggressively in the first-period. This model illustrates the important point that the link between contract markups and overall profit to the firm is tenuous at best when one considers the multiple stages of procurement. It is interesting to note that the policy recommendations following periodic reviews of defense industry profitability are often misguided because this fundamental point is not well recognized in the design of profit policy.

Although price is a primary concern of the government, it is clear that most procurement competitions are multidimensional. For example, contractors bid on both quality and price. In his contribution to this volume, Che generalizes the Bower-Osband model by allowing firms to make choices regarding the level of quality provided.[10] As in the previous work, Che uses a stylized model where the contract is initially awarded through competitive bidding and then is recontracted under the cost-based pricing that characterizes profit policy. However, Che allows designs to be variable and actively competed in the initial stage. Under such circumstances, Che finds that the beneficial role of cost-based pricing in handicapping low-cost firms is weakened, primarily because firms are able to inflate quality to earn higher profits in the subsequent cost-plus environment. Thus, such inflation has the effect of increasing the probability of winning the bid in the first stage as well as increasing profits in the second stage. In this model, raising the markup will handicap the low-cost firm as in the Bower-Osband framework but will also cause excessive design competition, thereby reducing the potential gains associated with the reduction in information rents.[11]

These results crucially depend on the assumption that the government is myopic, choosing the winner based on the quality of the system and the bid for the first period, without considering the implications of quality for cost levels (and, therefore, cost-plus payments) in subsequent periods. A related assumption is that the government cannot commit to a scoring rule that takes full account of these future costs. As Che notes, this assumption may have some strong basis in reality, given that program managers typically have short accountability spans because of job rotation and turnover. For example, the tenure of program managers is usually two to three years, even though procurements span 15 years.

---

[10]This paper is related to an article by Che (1993). This version emphasizes some of the institutional details and discusses policy implications. Those interested in the finer technical derivations should refer to that article.

[11]This gold-plating feature is related to the Averch-Johnson effect in the regulation literature.

Although the importance of endogenously determined quality in a cost-plus environment has been studied previously,[12] Che is the first to examine this issue using this more realistic regulatory setting. The main conclusion of this work is that, in theory, optimal contract design must come to grips with the necessity of considering both quality and price in an auctioning environment and, in particular, government decisionmakers must be able to commit to an evaluation methodology that considers the long-run cost ramifications of initial design decisions.

Even though the previous papers take the level of competition as given, a great deal of policy attention is directed toward enhancing competition in defense procurement. In his paper, Lars Stole considers alternative mechanisms for strengthening the role played by rival firms when incumbent developers enjoy significant cost advantages. As discussed above, the absence of complete information about contractor costs makes it difficult for the government to achieve the minimum possible price (the lowest cost plus a fair profit). The contractor can earn extra profits, or information rents, because the government cannot possibly audit true economic costs even after they have been incurred. One policy for reducing such rents is to use competition between contractors. As we have seen, such competition serves to award the contract to the most efficient producer but at a price roughly commensurate with the second-best bidder's break-even cost (true economic costs plus a fair rate of return).

In practice, however, there is often a significant wedge between the best and the next-best qualified firm. At the extreme, only one firm, often the original developer of the system design, may be qualified to bid. In such an instance, the information rents available are quite high and the government faces a large disadvantage in negotiating a reasonable price. Thus, the government is forced to consider alternative mechanisms for generating more competition.

Several such policy alternatives exist. These include dual sourcing, mandated teaming, and technology transfers. Dual sourcing guarantees the continued existence of multiple competitors (almost always two) by splitting production buys between firms who are bidding for higher shares of the total volume of the procurement. Teaming involves a mandated partnership by multiple firms in the early stages of development. Partners are required to split into separate competitive entities in subsequent stages following development. The third option, transferring technology to additional firms, is the focus of the paper by Stole.[13]

As Stole notes, technology transfers raise several policy questions that require attention. First, under what conditions should a second source be created? Second, should the second source be required to use the original design or, instead, can it use its own technology? Third, to what extent does transfer policy affect incentives in the initial development stage? As we have seen in the case of markups permitted by the

---

[12]See, for example, Rogerson (1990).

[13]Technology transfer can take three forms. First, the developer may be required to turn over a technical data package to the government. Alternatively, the original development contract may obligate the firm to transfer the necessary technology to a second firm. Finally, the terms of the development may contain predetermined fees for licensing the technology to another firm.

weighted guidelines, the contractual terms and conditions of a later stage of procurement can have implications for a firm's decisionmaking early on.

Imagine that there is a developer and a single potential second source. In Stole's model, each knows its own cost of production and can assess the likely range and associated probabilities of the rival's cost structure. The government can hire the developer or the second source with each employing its own technologies. Alternatively, the government can promote the transfer of the developer's technology to the second firm. In this case, the costs will be equal to the transfer costs (assumed to be a fixed amount) and production costs falling between the incumbent's and entrant's costs (that is, the transfer is not completely effective[14]). Clearly, with positive transfer costs and an inability to reproduce technology identically, total costs will never be minimized under this third option. However, if the threat of the transfer alters strategic bids by the firms, a policy that commits the government to a technology transfer under certain conditions might well be optimal in reducing expected payments for the procurement.

The government's problem, therefore, is to construct a decision rule that selects one of three options in response to bids provided by both firms. A conventional cost analysis might well conclude that the transfer option is merited if and only if the second source's bid plus the cost of transfer are less than the developer's bid. However, such a rule may result in inflated bids (above costs) by the incumbent. Thus, the optimal rule for choosing may not always minimize the government's payments for each and every set of bids. Rather, the goal of the rule is to diminish expected prices across a wide range of potential bids. One way the rule accomplishes this is by handicapping the developer's bid. Although the developer remains more efficient, the new entrant is more of a threat to win the bid, thereby forcing a strategic reduction in the average bid offered by the developer. This average price reduction justifies the occasional necessity of incurring transfer costs and switching to the less-efficient firm.

In general, the government will wish to commit to choosing the second source and mandating a technology transfer if the difference between the bids does not exceed some maximum threshold. This strategy will be desirable even if the developer's bid in certain instances remains below that of the second source (because of transfer costs and the inability to replicate the technology). Even though this policy might cost the government money in a particular instance, if information rents can be reduced via more aggressive bidding, this policy could reduce costs to the government on average.

Stole also points out that this licensing policy should be modified in instances when the developer's initial investment in technology is unobserved. For example, suppose one wishes to retain incentives to invest in quality. In such circumstances, it remains desirable to handicap the developer in competition with the second source.

---

[14]It is likely that the developed technology was tailored to the particular strengths of the incumbent. In addition, it is unlikely that the inexperienced rival will be unable to perform as well as the developer, either because the transfer is not completely effective or because they are not as far down the learning curve.

However, it would, at least in theory, be preferable to reduce the handicap in direct proportion to the quality of the final product.[15]

The next two articles in the volume directly address the issue of optimal contract design in the real-world context of defense procurement. Abstracting from quality considerations, the government's problem is to design a contract that minimizes total government expenditures over the entire procurement life cycle. As discussed above, the difficulty facing the government in designing such a contract is that it must account first for the fact that firms are better informed about their own costs and second, that they have discretion about how hard they work at cost reduction. Regarding the first issue, firms can more accurately predict costs before they are incurred and measure actual costs after the fact than can the government. Although government auditors can assess accounting costs, they are much less able to evaluate true economic costs that, for example, include opportunity costs, and thereby consider the level of effort, quality of labor, and the degree of top management involvement. In addition, in contrast to accounting measures, economic costs for decisionmaking should not include arbitrarily allocated overhead expenses that have nothing to do with increments to costs that are actually paid by the firm. Regarding the second issue, firms can pad costs by using less-efficient production techniques or inferior inputs, or simply not diligently lowering costs over time through experience.

Fixed-price contracts provide appropriate incentives for cost reduction, since firms will always make higher profits by lowering costs and the government will capture some of these expected cost reductions in the bidding process.[16] However, because the government cannot completely predict differences in firm costs, the low-cost firm will earn a potentially large information rent by bidding a price equal to (or slightly lower than) a less-efficient, higher-cost bidder. A cost-plus contract avoids this problem, but the government ends up paying a higher price than desired because a cost-plus contract gives no incentives for cost reduction—all cost savings are passed on to the government. DoD attempts to balance these problems by using a combination of the two contract types, called cost-plus-incentive fees where the firm bears a portion, usually between 15 and 30 percent, of the cost overruns.

In the contribution by Reichelstein, he first summarizes the literature on contracting theory in a sole-source, single-period environment. The theory provides a complete solution to the contracting problem. The optimal contract varies in an intuitive way, as a function of the likely cost dispersion among the bidding firms and their future opportunities to take action to reduce costs. Specifically, if the dispersion in costs is expected to be large, then usually cost-plus contracts to minimize potentially large information rents are favored, whereas if substantial cost-reduction opportunities are available, then fixed-price contracts to maximize the expected cost reductions are favored. An interesting twist of the theory is that the type of incentive contract

---

[15]As Stole points out, for an extremely high-quality design, it may actually be preferable to handicap the second source relative to the designer, depending on the initial cost distributions.

[16]Of course, this simple characterization ignores two important complications. First, firms could have incentives to reduce quality dimensions that are difficult to monitor. Second, if firms are risk averse and costs are unpredictable, even by them, they may be induced to bid a very high fixed price.

offered to the winning firm varies with its bid (and thus its cost); the government should offer a "menu" of incentive contracts.

The theoretically derived optimal menu of contracts can be recast in an intuitive format known as "variance-based schemes," which are similar to those used in procurement contracting. Under variance-based schemes, the incentive fee awarded to a firm should be determined by the variance between its forecasted and actual cost. Such schemes induce the contractor to forecast expected procurement costs accurately, thereby facilitating government planning. Perhaps more important, since these schemes are equivalent to the optimal contract, no other type of scheme can minimize total procurement cost more effectively. By offering a "menu" of contracts, designed so that each alternative will appeal to a firm with a particular cost structure, firms will reveal private information about costs, bid competitively, and induce the best possible outcome from the government's perspective.

Unfortunately, although the reviewed literature does quite well in characterizing "optimal" contract design, there remains a significant gap between theoretically pre-scribed mechanisms and what DoD is capable of doing in practice. So Reichelstein, in a study initiated by the German Department of Defense,[17] developed a con-structive procedure whose inputs are a set of project-specific parameters and whose final output is an optimal variance-based incentive scheme. The result is basically a refinement of cost-plus-incentive-fee contracts in which the choices differ by target profit and cost share. Not surprisingly, high target profit contracts carry low-cost shares, and vice versa. By submitting its cost estimate, the firm selects an "entrée" from the menu.

Reichelstein's contributions are twofold. First, he develops the theoretical point that variance-based incentive schemes are preferred to current DoD practice, at least in theory. Second, he demonstrates that they can be implemented in a mechanistic manner, without a detailed understanding of the underlying incentive theory. How-ever, even though the theoretical literature suggests that a menu of contracts should be optimal, one should not ignore the impracticality of constructing and im-plementing elaborate contract forms that require the design of auctions, menus, *ex post* audits, and a higher level of commitment to a particular type of contract over a long period of time. Thus, it is essential to balance the magnitude of the likely efficiency gains against the likely rise in administrative costs, even if Reichelstein's mechanistic procedure is adopted.

In the next contribution, Bower attempts to gauge the efficiency loss from the procurement designs actually employed by DoD. In practice, DoD offers a single cost-plus-incentive contract, not a menu of contracts from which the firm chooses.

---

[17]The author discusses several institutional restrictions placed upon the procedure by the German government, some of which are very similar to U.S. institutional restrictions. For example, the author discovered, in consultations with government officials, that cost ceilings had to be used in their procurement contracts. While cost ceilings may be incorporated into the procedures, the optimality properties of the incentive scheme may no longer hold.

Further, DoD and the contractor often renegotiate a contract at a future date.[18] Bower derives the optimal contract in a two-period model, but then conducts a series of simulations comparing optimal contract performance to cases where DoD renegotiates a contract that had previously been competitively awarded (so that DoD already has some information about actual costs) and where a single incentive contract is offered. He finds that the currently restrictive contract forms, if used correctly, achieve nearly optimal results. The gains to optimal contracts tend to be relatively meager, much less than those obtainable from introducing an additional qualified bidder within the existing contracting framework.

In particular, Bower demonstrates, at least for some specifications of the model, that savings from using a menu are rather small, amounting to as little as 0.2 percent of total procurement costs. However, the advantages of using a correctly calibrated incentive contract (even if there is only one incentive contract for all types of firms) rather than a simple fixed-price or cost-plus contract, are quite large, often over 10 percent. Therefore, the procedure developed by Reichelstein can be quite useful in suggesting cost share parameters that induce vigorous bidding, even if a menu of contracts is not offered. Finally, he also shows that the losses from renegotiation are small.

The final contribution, by Rogerson, models the independent actions of the Congress and the military. In the process, he provides one possible explanation for the observation that defense plants appear to be constructed inefficiently large given the ultimate volume of production. In Rogerson's model, the military, wishing to maximize defense capability, makes the initial decision to purchase plant capacity. The Congress reacts to this capacity, ordering the number of weapons that maximizes its own objective (specified as the difference between military benefits and costs). The argument made is that the Services, by building inefficiently large plant capacity, alter the marginal cost of producing weapons and, as a result, induce Congress to purchase a larger quantity than it otherwise would.[19] This outcome is made possible because Congress does not have the information needed to determine optimal scale (but, by assumption, the Services do) and because it does not have the power to commit in advance to a fixed military budget.[20]

Of course, alternative models can explain the existence of overly large plant scale. For example, Congress may not guarantee that unspent funds will not be recalled,

---

[18]In optimal contract design, long-term precommitment not to renegotiate a contract dominates renegotiation *ex ante* because it minimizes the opportunistic renegotiation by the contractor. This result is standard in the contracting literature, but a full explanation is beyond the scope of this introduction.

[19]One might argue that military weapons require fixed combinations of inputs, including materials, labor, and equipment. This is not likely to be strictly true, but it is plausible that there are limited opportunities for input substitution. This would diminish the likelihood that capacity distortions could alter marginal costs and thereby alter Congressional decisionmaking. Still, the posited mechanism could still operate in circumstances where future defense requirements are uncertain, and the Services wish to create inefficiently large excess capacity given reasonable expectations concerning demand surges. In the event of a surge, the incremental costs of additional output will be lower, thereby altering Congressional incentives on the margin.

[20]With additional Congressional oversight concerning *ex ante* costs and an increased willingness to negotiate all-or-nothing contracts, this distortion would be eliminated.

thereby inducing the Services (and contractors) to prefer spending money early in a procurement (i.e., on capacity). Or perhaps excess capacity is a rational choice made by contractors to reduce their marginal costs during future, more competitive procurements. Alternatively, perhaps the large scale is a decision on the part of both Congress and the military to invest in the defense industrial base in preparation for uncertain surge requirements. Unfortunately, these alternative explanations cannot be distinguished empirically from Rogerson's formulation.[21]

Regardless, this paper makes two interesting contributions to the literature on defense procurement. First, it represents a refreshing departure from traditional analyses of the government-contractor relationship. It demonstrates that by explicitly considering the multiple and often conflicting objectives of different government agencies (in this case Congress and the military), economic models can explain many of the stylized facts of defense procurement that may appear puzzling in the context of a single procurement agent. In addition, this paper draws attention to potentially important efficiency losses associated with cost-based pricing. That is, if prices are based on audited costs, a monopolist will have an incentive to distort input choices to reduce marginal costs in favor of higher fixed or sunk costs, much in the same way that a private sector monopolist will incur transactions costs to implement a two-part pricing scheme. Indeed, there are alternative mechanisms for achieving the same outcome that are also worthy of further study. For example, the Services can use indemnification clauses, requiring lump-sum payments to contractors if the Congress cancels a program. Also, the military can invest heavily in production processes that require substantial up-front costs while reducing marginal costs. Such investments may not be cost effective from the Congressional perspective. Finally, the military may overinvest in product quality, incurring high fixed costs but altering the subsequent decisionmaking calculus by increasing the marginal value of additional quantities.

4. POLICY IMPLICATIONS

The analytic efforts presented and cited in this volume attempt to narrow the wedge between economic theory and the real-world markets, institutions, and regulatory environments prevailing in defense procurement. The reported models explicitly consider various aspects of this complex process, including the structure of profit policy, the multiple stages of defense contracting, interrelationships between contemporaneous contracts, the absence of adequate information for decisionmaking, and independent roles played by market participants with inconsistent objectives. As a result, these models provide fresh insights that should be considered in the design of policies meant to facilitate competition, select the right contractor, and provide appropriate incentives to produce high-quality weapons at the lowest cost. Some of the more interesting policy implications include:

---

[21] In fact, one might argue that these assumptions are more plausible than Rogerson's, which relies on an extremely myopic Congress. In reality, plant capacities are often built at Congressionally approved levels, and production is then reduced in subsequent budget crunches.

- The structure of the weighted guidelines inadequately compensates firms for unrecognized opportunity costs and for categories of risk.

- Rules regarding the allocation of overhead distort production decisions. For example, on cost-plus contracts, firms are likely to use excess labor and overly rely on subcontractors. On fixed-price and commercial contracts, firms will overinvest in automation. Where possible, costs should be allocated directly rather than via overhead pools.

- In the long run, profits available on particular contracts will be dissipated by competition to win the contracts. For example, profit policy is more likely to affect R&D and innovation than contractor profits overall.

- Since contracts are serially related, the weighted guidelines could have counterintuitive effects on firm profitability. For example, over a wide range, higher markups after contract renegotiation or on follow-on procurements can stimulate more aggressive bidding by firms during the initial competitive stages that more than compensates for the higher payments later.

- If program personnel are myopic (because of short tours of service, for example), contractors will overinvest in quality, taking advantage of the cost-plus nature of pricing in later stages of procurement. This suggests that evaluation criteria for awarding contracts should handicap high-quality providers, at least to some degree.

- Technology transfers, even when seemingly inefficient from a single contract perspective (i.e., the transfer is expensive and the technology is imperfectly adopted by the second source), may be a desirable policy option. This is because the threat of dual sourcing will induce more aggressive competition and, overall, will create cost savings for the government. The decision to transfer technology should depend on the firms' bids and the government's prior knowledge of contractor costs.

- Optimal contracts can be approximated by simple algorithms that flexibly choose cost share parameters as a function of contractor characteristics. If such parameters are chosen correctly, the savings to the government can be considerable. However, the efficiency gains associated with the more complex auctions, menus, and full-commitment contracts suggested by economic theory are unlikely to produce significant improvements to warrant the increased difficulty of implementation.

- Because of the contrasting incentives of different government agents in the procurement process (such as Congress versus the Services) along with the cost-based nature of pricing, certain avoidable inefficiencies will emerge. For example, the military will have incentives to reduce variable costs at the expense of the fixed or sunk costs that do not influence future decisionmaking by the Congress.

PROFIT IN NEGOTIATED DEFENSE CONTRACTS: A SURVEY OF SOME RECENT LITERATURE
by John H. Nachbar

1. INTRODUCTION

The U.S. Department of Defense (DoD) negotiates procurement contracts primarily within a framework known as profit policy. Briefly, profit policy consists of a list of possible contract forms and a complex set of guidelines governing the level of "profit" contracting firms may be allowed to receive. The standard example of a negotiated contract is for the production of a highly sophisticated weapon. In such cases, the original developer is arguably the only qualified producer and so there is no scope for awarding the contract by price competition (e.g., via an auction). However, the influence of profit policy is more pervasive than this example suggests. Even competitively awarded contracts may employ profit policy if the contract is renegotiated at midterm, say because of a design change. Moreover, in "negotiated competitive" auctions, which have seen increasing use, profit policy is employed to set boundaries for auctioned contracts.

The total fraction of DoD contracts awarded under profit policy is believed to be very large, but definitive numbers are elusive because of data and definitional problems. For some idea of the magnitudes involved, in fiscal 1983, negotiated awards totaled $106 billion, about three-quarters of total DoD awards. About one-quarter of the $106 billion were "negotiated competitive" contracts. Schmidt (1993) estimates that by 1988, the total fraction of negotiated awards had declined, but still accounted for roughly 60 percent of total awards. Of these, roughly half were "negotiated competitive." The increased use of negotiated competitive contracts, from 19 percent of the total in 1983 to 30 percent in 1980, may reflect implementation of the 1986 Competition in Contracting Act, aimed at discouraging sole-source, noncompetitive awards.

This paper provides an overview of some recent economic research on the structure of profit policy.[1] The work surveyed, Osband (1989a, 1990a) and Rogerson (1992a), asks whether profit policy is well formulated to exploit profit as a policy instrument, on which see Section 2. Their answer is "not completely." Much of the contract

---

[1] The research focuses on production contracts but much of the analysis carries over to contracts for R&D or services.

profit under profit policy is repayment for current expenditures not reimbursed more directly, for example, interest charges on working capital. What will be called "true" profit, payment in excess of what is needed to induce a firm to accept a production contract, is disguised as a residual. Moreover, some of the formulas used by profit policy to compute cost reimbursement may be flawed. The practical importance of this conclusion is unclear. DoD negotiators have substantial latitude and can, in principle, compensate for deficiencies in formal profit policy guidelines. However, the finding that profit policy guidelines are neither operationally transparent nor, in many respects, sensibly structured may be reason enough to justify revision.

The remainder of the survey is organized as follows. Section 2 discusses the economic role of profit policy and offers some perspective on the relationship of the research surveyed here to other profit policy research. Section 3 provides a brief overview of profit policy structure. The presentation is essentially an abbreviated version of the survey provided in Rogerson (1992a). Additional detail can be found in Osband (1989a and 1990a). Section 4, the heart of this survey, computes "minimum profit" in a simple case. The structure of the analysis is based on Rogerson (1992a) but many details are drawn from Osband (1990a). Section 5 compares the solution found in Section 4 with actual profit policy as sketched in Section 3. Finally, Section 6 contains some remarks on where further research may be most beneficial.

2. PROFIT AND PROFIT POLICY

True profit influences long-run patterns of industry structure by serving as a "prize," inducing firms to enter the defense industry and invest in facilities capital, in independent research and development, and in bids and proposals. This is emphasized in Rogerson (1992a); see also Lichtenberg (1988a), and Tan (1989). In long-run equilibrium, total expected profits will be exhausted, or nearly exhausted, by the expenditures necessary to win a contract. Thus, profit ultimately influences not the size of defense industry profits, which ought to be zero after correcting for the cost of capital, but rather the size and structure of the "defense industrial base."

One may question how effective profit can be as a policy tool. There are, first, subtleties in how profit policy is perceived by firms. Firms contemplating an investment may fear that future DoD policy could favor competitive contracts over negotiated contracts, e.g., the 1986 Competition in Contracting Act. For theoretical perspective on this problem, see Laffont and Tirole (1988a) and Riordan and Sappington (1989). Firms may also fear that the levels of profit offered by production contracts at the time they make their investment will not persist, that profit policy may become less generous; see, for example, Tirole (1986). Moreover, even if DoD could commit to a stable profit policy, the link between contract profit and overall profit to the firm might not be straightforward. For an example due to Bower and Osband (Chapter Four), consider the fairly common circumstance in which profit policy is used in the renegotiation of contracts that were originally competitively awarded. Because contract profit under profit policy is based on markup over cost, firms whose costs are intrinsically higher get higher profits, in absolute terms. As a consequence, in the

initial competitive bidding phase, high-cost firms bid relatively aggressively. The lowest-cost firm still wins the auction, but its overall profits, combining the first and second phases of production, are squeezed. The model suggests that in some circumstances the operation of contract profit may be the opposite of what is expected: The more "generous" is profit policy in awarding profit markups, the less profit is actually transferred to firms and therefore the less investment is induced.

Profit is also hampered as a policy tool by its lack of precision: Overall profit rewards not only "legitimate" precontract expenditures but also outlays that may be unwanted. Lobbying is a frequently cited example of socially wasteful investment. As another example, Osband (1990b) argues that, again because profit policy computes profit as a markup over cost, firms may be induced to develop equipment that is more sophisticated, hence costly, than is socially desirable.[2]

These problems noted, it may be argued that more direct methods of reimbursement need be no more efficient. To provide incentive for hard work, reward for research programs should depend at least in part on the success of the research. However, DoD may lack expertise in evaluating research, especially in its early phases, and research teams may also fear that DoD, to avoid paying a bonus for good work, has an incentive to play down the importance of research breakthroughs. Profit policy finesses these problems to a degree by, in effect, making production the criterion for research success. Firms are left to their own judgment as to what research to pursue, and their decisions are disciplined by the knowledge that their own funds are at stake.

Quite apart from the question of the "optimum" level of contract profit, there are important issues concerning the structure of the contracts themselves. These matters lie outside the scope of the research upon which this survey focuses, but for completeness some of the work in this area will be described briefly.

A fundamental difficulty confronting the designer of a DoD production contract is that a firm will typically be better informed than the government both about its own cost structure and about the efforts it has taken to make production efficient. This raises the possibility that, on the one hand, the government might unknowingly reimburse a firm for costs not actually incurred, and on the other, that the firm may have insufficient incentive to keep costs under control. Taken individually, these problems admit straightforward solutions. The first can be addressed by reimbursing only audited actual cost. The second can be handled by reimbursing only expected cost, so that the firm bears the full burden of cost overruns (and the full benefit of cost underruns).

One may note, however, that these two solutions conflict. It is, in fact, impossible to solve both problems simultaneously. The message of the literature on optimal contracts has been that, rather than attempt a complete correction of just one problem, it is better for the government to make partial progress against both. The scheme

---

[2]See also Che (1993). The argument is delicate because the government has considerable discretion in awarding contracts and in assigning markups.

that implements this approach has the government offer potential producers a "menu" of different contracts from which they can choose. The contracts are incentive contracts: Firms bear at least part of any cost overrun, and conversely, they benefit from at least part of any cost underrun. The menus are constructed in such a way that a firm reveals its cost structure by its menu choice. When there are several qualified producers, use of menus enables the government to identify and select the producer with lowest cost. However, to get firms to reveal their costs truthfully, the government must pay an "information rent" to the producing firm. Moreover, this construction necessitates that the government settle for a second-best level of cost-control effort. For more on this line of work, see Laffont and Tirole (1986, 1987), Baron and Besanko (1987b), and Reichelstein (Chapter Seven), and see also the review of the regulation literature in Caillaud et al. (1988).

The contracts actually employed under profit policy closely resemble full menu contracts. Menuing may not be as elaborate as the theory specifies but it is conceivable that DoD contract forms are optimal, once the cost of constructing elaborate menus is taken into account. Some progress toward gauging the efficiency loss from actual DoD contracting is made in Bower (Chapter Eight). Numerical calculations, though merely illustrative, suggest that efficiency losses from simple profit-policy-type contracts relative to full menu-type contracts may be small, provided the profit policy contracts are otherwise correctly designed.

3. AN OVERVIEW OF PROFIT POLICY STRUCTURE

It is useful to think of profit policy as being composed of two somewhat independent components: contract forms and the weighted guidelines.

Contract Forms

Let C denote total recognized project costs. Recognized costs include direct material and labor costs and many categories of overhead costs. Not included are expenditures on independent research and development, bid and proposal expenses, the rent on facilities (land, buildings, and equipment; if the firm is the owner of the facilities, the rent is imputed), general and administrative expenses (G&A; essentially central management overhead), and interest charges on working capital.[3] All of these expenditures are compensated through contract profit, if they are compensated at all.

Under the canonical contract form, the government pays the firm, upon project completion, p:

$$p = (1 - \alpha)C + \pi^* - \gamma(C - C^e) \tag{3.1}$$

---

[3]Working capital is net short-term liabilities. Such liabilities might be incurred for the purchase of raw materials or the payment of wages, if the government does not provide immediate reimbursement. Until reimbursement is paid, the working capital incurs interest charges.

where α is the progress payment rate, π^* is contract profit, γ is a cost-sharing parameter, and C^e is the cost expected at the time the contract is signed. Progress payments are reimbursements paid at, or shortly after, the time costs are incurred. At date t, if cost C_t is incurred, then the government pays immediately an amount αC_t. Thus, by the end of the contract, $\sum \alpha C_t = \alpha \sum C_t = \alpha C$ have already been reimbursed via progress payments, leaving an additional $(1 - \alpha)C$ to be paid at contract completion. Progress payments are modeled here as applying only to cost, which accords with what are called fixed-price contracts. By comparison, under cost-plus contracts, α is set equal to 1 and applies to profit as well as cost. Regarding cost sharing, if the firm incurs a cost overrun $C > C^e$, then the firm is "taxed" $\gamma(C - C^e)$. Note that this tax could be negative (a subsidy) if the firm instead achieves a cost underrun. This gives the firm incentive to keep actual costs C low. Ignoring questions of payment timing and working capital costs, if $\gamma = 0$ then the government reimburses actual costs C in full (full cost sharing); if $\gamma = 1$, the government pays only C^e (zero cost sharing). Contract types are surveyed in somewhat greater detail in Appendix A. The discussion in Section 4 will focus primarily on fixed-price-type contracts, although it can easily be extended to cost-plus contracts.

Contract profit, π^*, is determined by the weighted guidelines, described below. It will be assumed throughout that C^e is "correct": Before signing a contract, the government and the firm are equally well informed and agree on the distribution of possible outcomes. In fact, it is likely that the firm and the government are not equally well informed, but this complication is not considered in the work surveyed here.

The Weighted Guidelines

The weighted guidelines specify π^* as a function of a number of variables. These are summarized in Table 2.1, which is taken with only slight modification from Table 2.4 in Rogerson (1992a). The entries will be discussed briefly in turn.

Working Capital. As discussed in Appendix A, cost-plus contracts receive 100 percent progress payments and therefore involve essentially no working capital. It is therefore appropriate that such contracts receive no markup for working capital. A fixed-price contract of length n receives a markup equal to the product of $\tau L(n)$, where τ is the "treasury rate" and L is a time factor (see below), and $(1 - \alpha)C^e$, which is the fraction of expected cost that is not reimbursed immediately through progress payments.[4] The treasury rate, sometimes also referred to as the CAS 414 rate, is an estimate by the U.S. Treasury of the current rate on five-year commercial loans. It has been examined by Rogerson and found to be closely approximated by the rate on medium-term (five-year) U.S. government bonds plus one percentage point.

---

[4]Fixed-price contracts that pay no progress payments employ a different compensation scheme for working capital. This alternative scheme is seriously flawed. For example, the effective interest rate is fixed at 2 percent. Perhaps partly as a consequence, progress payments are paid on all major contracts.

Table 2.1

The Weighted Guidelines

Component of Profit	Normal Value, %	Allowable Range, %	Base to Which Percentage Is Applied
Working capital			
Cost-type contracts	—	—	—
Fixed-price contracts with no progress payments	2.0	2.0	C^e
Fixed-price contracts with progress payments	Treasury rate	Treasury rate	$\hat{L}(n)(1-\alpha)C^e$
Facilities capital			
All contracts receive	Treasury rate	Treasury rate	nK
Standard extra profit			
Land	0	0	nK_L
Buildings	15	10 to 20	nK_B
Capital	35	20 to 50	nK_E
Alternate extra profit	2	2	C^e
Performance risk			
Technical risk	1.2	0.6 to 1.8	C^e
Management	1.2	0.6 to 1.8	C^e
Cost control	1.6	0.8 to 2.4	C^e
Total	4.0	2.0 to 6.0	C^e
Contract risk			
Firm fixed price (FFP)	3.0	2 to 4	C^e
Fixed plus incentive (FPI) or cost plus incentive fee (CPIF)	1.0	0 to 2	C^e
CPFF	0.5	0 to 1	C^e

Up to a maximum of 2.9 at 76 months or more, $L(n)$ is a step-wise approximation to the piece-wise linear function:

$$\hat{L}(n) = \max\{0, n/2 - 0.35\}$$

where n is contract length. It will be argued in the next section that $1/2rn(1 - \alpha)\ C^e$ yields a possible approximation of the "correct" working capital markup when (a) there is no risk and (b) cost is expected to be incurred uniformly over time.[5]

Facilities Capital. Facilities capital compensation consists of two components. The first, the facilities capital cost of money, is an interest charge on the product of the length of the project, n, and the book value of the facilities capital employed in the project, K.[6] The interest rate used is again the treasury rate. The second component, the facilities capital markup, is specific to the category of capital employed, land (L), buildings (B), or equipment (E).[7]

---

[5]The rate r is the interest rate on a "riskless" investment (usually taken to be a government treasury bill).

[6]This component is, technically, not part of the weighted guidelines.

[7]The guidelines also provide for an alternative to the facilities capital markup. This is typically employed for projects, such as "pure" research and development projects, employing little in the way of facilities capital. As Rogerson notes, this is on its face bizarre: Projects with no facilities capital should receive no facilities capital markup. Conceivably, the markup is intended to compensate for some other unrecognized cost, but if so, as Rogerson argues, this should be made explicit.

Performance Risk. The natural interpretation of technical risk is that it is compensation for uncertainty surrounding cost. If the firm shares at least some of the cost variability, $\gamma > 0$, then the firm's owners will require compensation for bearing the associated risk. As can be seen from Table 2.1, profit for technical risk is computed as a function of expected cost only. The computation ignores the cost-sharing rule, γ, as well as other parameters that might conceivably matter, such as the length of the contract. Rogerson suggests as one partial explanation for this that technical risk may be intended to provide compensation for certain unrecognized costs other than risk. It may, for example, be that certain projects absorb a larger fraction of the time of the company's best managers and engineers than is acknowledged under standard accounting rules. It is possible that these overhead-intensive projects are also riskier, and hence a contract might write compensation for such unrecognized costs as a function of risk if risk were relatively easy to observe and verify.

As discussed at length in Rogerson (1992a), the other entries under "performance risk" are not compensation for risk either. "Cost control" is a reward for the contractor's past excellence, as perceived by the government, in minimizing cost. As Rogerson (1992a) notes, the motivation for this component is not entirely clear. If cost control is observable after contract completion, then one would think it would be possible to provide a reward for good control, and a penalty for bad control, in the current contract, rather than postponing the incentive to the firm's next contract. "Management" covers a variety of concerns. Partly, it is a reward for perceptions that management has been effective in some sense; to a degree this presumably overlaps with cost control. Partly also, it is intended to provide extra compensation for those projects that demand greater management attention than accounted for in C.

Contract Risk. Contract risk provides compensation for cost uncertainty. Unlike the technical risk correction, contract risk depends on contract type, this providing a crude link to the cost share γ; see Appendix A. There is no explicit dependence on contract length.

4. A THEORY OF CONTRACT PROFIT

The derivation here, based largely on Rogerson (1992a) with some modifications due to Osband (1990a), is meant to highlight the structure of the theoretical argument rather than to offer a finished product for computing contract profit. As one simplification, some components of unrecognized costs, for example, G&A expenses, will be ignored. Thus, it will be assumed that C comprises all current project costs save working capital expenses and facilities capital rent. It will also be assumed that total expected cost, C^e, is incurred uniformly over time:

$$C_t^e = C^e / n$$

where, again, n is the length of the contract. For simplicity, taxes, which are incorporated in Osband (1989a), will be ignored.

For clarity, the discussion will first handle the case where there is no cost uncertainty, thus, expected cost C_t^e equals actual cost C in every period and there is no role for risk compensation. The facilities capital component of contract profit will be derived only for the no cost uncertainty case.

No Cost Uncertainty

Working Capital. The task is to find the minimum contract profit π_w such that the payment at the conclusion of the contract just compensates the firm for its working capital costs (zero true profit). Ignoring facilities capital, from Eq. (3.1) the minimum contract payment is given by:

$$p = (1-\alpha)C + \pi_w \tag{4.1}$$

where use has been made of the fact that $C = C^e$. The term $(1-\alpha)C$ compensates the firm only for the principal on working capital. Interest on working capital must therefore be covered by contract profit. Interest charges accrue at a rate r per period whether the funds are borrowed internally or externally.[8]

Computation of π_w requires that the cost stream $\{-(1-\alpha)C/n\}$ and the payment p be placed on a comparable financial basis. This will be done by discounting all costs and payments to the start of the contract, date $t = 0$. Recall that a date 0 amount V is said to be financially equivalent to $\{-(1-\alpha)C/n\}$ if V, deposited at $t = 0$ and earning the interest rate r, exactly covers withdrawals of $\{-(1-\alpha)C/n\}$ per period for n periods. Thus, paying V today or $\{-(1-\alpha)C/n\}$ for n periods are equally costly to the firm. The V that will do this is given by:

$$V = -\frac{(1-\alpha)C}{n}\frac{1}{(1+r)} \cdots - \frac{(1-\alpha)C}{n}\frac{1}{(1+r)^n}. \tag{4.2}$$

Similarly, the amount W which, if deposited today at the interest rate r, would yield exactly p in n periods is given by:

$$W = p\frac{1}{(1+r)^n}. \tag{4.3}$$

Thus, the firm is indifferent to receiving W at $t = 0$ or receiving p at $t = 1$. To determine the minimum contract profit π_w, we want the contract to involve zero true profit, hence we want $V + W = 0$. This yields:

---

[8]Throughout, r will be used to denote the "riskless" rate. A plausible real world candidate for r is the interest rate on short-term government debt, in particular treasury bills (t-bills). In practice, t-bill rates fluctuate over time, as inflation varies if for no other reason, but this fluctuation will be assumed away. In the absence of any uncertainty, as here, r is the unique interest rate. In the presence of uncertainty, there will be many expected interest rates, varying according to risk.

$$p = \frac{(1-\alpha)C}{n} \sum_{t=1}^{n} (1+r)^{t-1}.$$ (4.4)

Substituting from Eq. (4.1) we have:

$$(1-\alpha)C + \pi_w = \frac{(1-\alpha)C}{n} \sum_{t=1}^{n} (1+r)^{t-1}$$

or

$$\pi_w = \frac{(1-\alpha)C}{n} \left[\sum_{t=1}^{n} (1+r)^{t-1} - n \right].$$ (4.5)

For future reference, let:

$$F = \sum_{t=1}^{n} (1+r)^{t-1} - n.$$ (4.6)

Again, π_w gives the minimum contract profit needed to cover the interest charges on working capital. If working capital charges are the only unreimbursed expense, then true profit is the difference:

$$\pi = \pi^* - \pi_w.$$

As this suggests, if $\pi^* < \pi_w$ the firm loses money by taking the contract.

The formula for working capital in the weighted guidelines can be viewed as a first-order approximation to Eq. (4.5), or rather to the analog of Eq. (4.5) in continuous time. The latter is given by:

$$\pi_w = (1-\alpha)\frac{C}{n} \left[\int_0^n e^{rt} dt - n \right]$$

$$= (1-\alpha)\frac{C}{rn} [e^{rn} - 1 - rn].$$ (4.7)

A first-order Taylor expansion around $n = 0$ (see, for example, Rudin, 1964) yields:

$$\pi_w \approx r\frac{(1-\alpha)C_n}{2}.$$ (4.8)

This approximates Eq. (4.5) by charging uncompounded interest on the "average" cost over the period, namely, $1/2(1-\alpha)\,C_n$. The reader will note that Eq. (4.8) differs from the working capital correction in the weighted guidelines in two respects. First, the "risk-free" interest rate r typically lies a percentage point below the treasury rate τ

used in the guidelines. Second, the time correction here is $n/2$, whereas the working capital adjustment is, essentially, $n/2 - 0.35$.

Facilities Capital. The derivation sketched here for minimal contract profit for facilities capital (e.g., land, plant, and equipment) most closely resembles that in Osband (1990a). The treatment will ignore many real-world complications, notably depreciation and taxes. We model a firm as having to choose the size of its facilities each period *before* the firm knows whether it will in fact win a production contract. For example, the firm may have to have substantial production facilities already in place to be considered a "qualified" bidder. Having made an investment of $\$K$ worth of facilities, the firm faces the possibility that it may not win a contract and that the facilities may remain idle. Formally, in each period there is a probability η that the firm will be awarded a contract, which if won lasts for n periods. We expect that η will depend on demand conditions and on how readily the facility can be put to different use (its "fungibility"). Typically, η will vary over time, but for simplicity we assume here that it is constant. Moreover, we assume that the uncertainty introduced by η has zero systematic risk and there is no other source of uncertainty.[9]

If the firm does not win the contract, the clock moves forward one period. For simplicity, a period is taken to be one year, but the derivation is easily modified to allow for periods of other lengths. By assumption, capital does not depreciate and can be resold on the open market at the beginning of each period. Thus the value of the capital at date 1, and at date n, should simply be $\$K$.[10] If the firm invests so that the value of capital exactly equals its cost to the firm, we should have:

$$K = \eta (\pi_{fac} + K) \frac{1}{(1+r)^n} + (1-\eta) \frac{K}{1+r} \tag{4.9}$$

where π_{fac} is the minimum value of the facilities component of contract profit. Solving for π_{fac}:

$$\pi_{fac} = \left[(1+r)^{n-1} \frac{(r+\eta)}{\eta} \right] K. \tag{4.10}$$

As η goes to zero, π_{fac} must go to infinity to compensate the firm for a low probability of being employed.

The government in principle has the ability to exploit the fact that, after facilities investments are made at the beginning of a period, these investments are "sunk" for the remainder of the period. To be concrete, suppose that a firm invests $\$K$ in facilities capital thinking that contract profit will be π_{fac}. However, suppose that be-

---

[9]Zero systematic risk here means zero covariance with the return on the "market portfolio." This is discussed in the subsection on cost uncertainty.

[10]This assumes the absence of a "bubble" in the market for facilities capital. That is, implicitly, a transversality condition is imposed here.

tween the time of the investment and the time of actual contract bidding the government changes the regulations so that contract profit is only $\pi' < \pi_{fac}$.

If the firm participates in the bidding and wins the project, it has essentially two choices. First, it can refuse to take the contract. In this case, its capital will have a resale value of $\$K$ at the start of the next period. The present value of this is:

$$\frac{K}{1+r}.$$

Alternatively, the firm can take the contract. The present value of doing so is:

$$(\pi' + K)\frac{1}{(1+r)^n}.$$

The firm will be just indifferent between these two options if:

$$\frac{K}{1+r} = (\pi' + K)\frac{1}{(1+r)^n}.$$

Thus, a winning firm will accept a contract for a profit of only:

$$\pi' = K[(1+r)^{n-1} - 1]. \tag{4.11}$$

π' will always be smaller than π_{fac}. Of course, if the government does exploit firms in this way, and firms learn to anticipate this exploitation, firms will not invest in facilities capital before bidding. This may harm DoD interests more than any short-term gains from substituting π' for π_{fac}.

Two other points may be noted. First, incorporating depreciation will increase π_{fac} and will also cause π_{fac} to vary by type of facilities capital (e.g., machine tools vs. land). π_{fac} will also vary by capital type because of differences in η stemming from differences in fungibility. Explicitly, for highly fungible investments, η will be high (close to 1) so that π_{fac} will be relatively low. Second, the randomness that matters here resides not in the cost stream but in whether capital will be hired at all. Although these two types of randomness may be correlated, conceptually it is useful to keep them distinct, a point emphasized in Rogerson (1992a).

Cost Uncertainty

Uncertainty will be modeled here as affecting only the cost stream $\{-\tilde{C}_t\}$, now a stochastic process. Other contract attributes, such as length, may likewise be uncertain. In addition, there are risks to the firm associated with the fact that production contracts are not complete (do not specify actions under every possible contingency) and with the fact that there are constraints on DoD's ability to make binding contract commitments. For example, the firm may be concerned that DoD may cancel the

project, or, following a design change, may force renegotiation of the contract at midstream, possibly in a changed regulatory environment. Profit policy analysis has not yet incorporated such uncertainty into the theory. This subsection will ignore facilities capital.

With uncertain costs $\tilde{C}_t$, Eq. (4.1) becomes:

$$p = (1-\alpha)\tilde{C} + \pi^* - \gamma(\tilde{C} - C^e) \tag{4.12}$$

which is random. Before giving a formal analysis, it is worth considering what kind of result might be expected. The firm will still need to be compensated for interest charges on working capital. In addition, we now expect compensation for risk. We do not distinguish, as the weighted guidelines do, between "technical risk" and "contract risk"—it is not clear what that distinction means. However, there *are* two different sources of risk to be considered. The first is that p is random. The second is that working capital interest payments are now also random. However, it will be shown that working capital risk is relatively unimportant.

The role played by these two risk components can be brought into sharper focus by following a trick due to Rogerson (1992a). Decompose the contract into two artificial contracts—a pure cost-sharing contract in which there is no working capital ($\alpha = 1$) and a pure financing contract in which there is no cost sharing ($\gamma = 0$). The two contracts yield, respectively, minimum contract profits of π_c to reimburse for cost-sharing risk, and π_w to reimburse for working capital interest charges and working capital risk. Since, as the reader can verify, the actual contract is equivalent to the sum of these two contracts, we have a minimum overall contract profit of:

$$\pi_m = \pi_c + \pi_w.$$

Very loosely, the expected value of the cost overrun $(\tilde{C} - C^e)$ is zero (by the assumption that C^e is the true expected value of $\tilde{C}$) but the present value of the cost share $-\gamma(\tilde{C} - C^e)$ may nevertheless be negative because of the uncertainty surrounding $\tilde{C}$.[11] Consequently, π_c will be positive provided γ is positive and should be maximal, all else equal, when γ is 1. It seems plausible that the minimum contract profit will be of the form $\pi_c = \omega_c \gamma C^e$; we expect that ω_c will vary with specifics of the contract. For the pure financing contract, meanwhile, principal is again repaid via $(1 - \alpha)\tilde{C}$, which, since it is equal to actual principal, requires no risk correction. Interest, on the other hand, requires risk correction because it is compensated by π_w, which is fixed even though $\tilde{C}_t$, and hence interest on $(1-\alpha)\tilde{C}_t$, is variable. A plausible formulation is $\pi_w = \omega_w(1-\alpha)C^e$, with ω_w a risk-corrected variant of F/n; see Eqs. (4.5) and (4.6). Again, we expect ω_w will vary with contract parameters. Summing up, we expect that:

$$\pi_m = [\omega_c \gamma + \omega_w(1 - \alpha)]C^e. \tag{4.13}$$

---

[11]A sufficient condition for this claim to hold will be provided below.

Although the intuition behind Eq. (4.13) is clear, providing a formal derivation is difficult. In Rogerson (1992a), and in subsequent work by Osband (1990a), V and W, and hence ω_c and ω_w, are computed via a theory of asset valuation known as the period-by-period capital asset pricing model (CAPM). The period-by-period CAPM is common in applied work but the assumptions that underlie it are strong. Sufficient conditions for the model to hold are discussed in Chamberlin (1988) and in Huang and Litzenberger (1988, Section 7.22). Both derivations, incidentally, hold exactly only in continuous time. See also Fama (1977).

For the derivation here, $\tilde{C}_t$ will be modeled as a stock variable; it is the sum of the date 0 cost level and all input price shocks and production innovations that have occurred since. Following Osband (1990a), it will be assumed further that the value of $\tilde{C}_t$ is independent of the order in which these shocks occur. This assumption is made somewhat more precise in Appendix B. There it is derived, in discrete time, that:

$$\omega_c = \frac{1}{n}\left[\sum_{t=1}^{n}(1+\mu)^t - n\right].\qquad(4.14)$$

$$\omega_w = \frac{1}{n}\left[(1+r)^n\sum_{t=1}^{n}\left[\frac{1+\mu}{1+r}\right]^t - \sum_{t=1}^{n}(1+\mu)^t\right].\qquad(4.15)$$

The term μ, a constant, is the risk coefficient:

$$\mu = -\frac{(\rho^e_{M_t} - r)}{\text{var}(\tilde{\rho}_{M_t})}\,\text{cov}\left(\frac{\tilde{C}_t}{C_t^e}, \tilde{\rho}_{M_t}\right)$$

where $\tilde{\rho}_{M_t}$ is the return on the "market portfolio" in date t, var denotes variance, and cov denotes covariance. The market portfolio is the collection of all capital assets, including not only all equity but also private real estate and human capital (e.g., the market value of education). In practice, the market is often taken to be a well-diversified portfolio of stocks. As in Rogerson (1992a) and Osband (1990a), it is assumed that

$$\text{cov}(\tilde{C}_t, \tilde{\rho}_{M_t}) < 0$$

hence $\mu > 0$. That is, roughly, the stock market falls when defense input prices rise. As both authors note, this assumption is not unexceptionable, but it is at least plausible. The assumption implies that the present value of the cost share $-\gamma(\tilde{C} - C^e)$ is indeed negative, hence π_c is positive. Appendix B provides some intuition as to why μ takes the above form.

As a check on Eq. (4.15), note that it yields Eq. (4.5) if $\mu = 0$.[12] The complexity of Eq. (4.15) obscures the fact that the influence of μ on ω_w is slight. This can be seen more transparently if we switch to continuous time. For completeness, the analogs of both Eq. (4.14) and Eq. (4.15) are given:

$$\omega_c = \frac{1}{n}\left[\int_0^n e^{\mu t}dt - n\right]$$
$$= \frac{1}{\mu n}[e^{\mu n} - 1 - \mu n] \tag{4.16}$$

and

$$\omega_w = \frac{1}{n}\left[e^{rn}\int_0^n e^{(\mu - r)t}dt - \int_0^n e^{\mu t}dt\right]$$
$$= \frac{1}{n}\left[\frac{e^{rn}}{(\mu - r)}[e^{(\mu - r)n} - 1] - \frac{1}{\mu}[e^{\mu n} - 1]\right]$$
$$= \frac{1}{n(\mu - r)}[e^{\mu n} - e^{rn}] - \frac{1}{n\mu}[e^{\mu n} - 1]. \tag{4.17}$$

Computing a first-order approximation, as for Eq. (4.8), we get:

$$\omega_c \approx \frac{1}{2}\mu n. \tag{4.18}$$

$$\omega_w \approx \frac{1}{2}rn. \tag{4.19}$$

The latter is exactly the coefficient in Eq. (4.8). Roughly, the intuition is that the risk term in Eq. (4.17) is approximately $\mu rn/2$, which will be quite small if, as is typically the case, both μ and r are of the order $1/10$ or less. Numerical simulation confirms this. For $n = 3$ and $r = 0.1$, ω_w as given by Eq. (4.17) rises from 0.166 at $\mu = 0$ only to 0.175 at $\mu = 0.05$. The approximation given by Eq. (4.19) is roughly 10 percent too low at $\mu = 0$, and this worsens to 14 percent too low at $\mu = 0.05$. See Figure 2.1.

True Profit and Other Compensation

The derivations above are illustrative. A more complete formulation would take into account G&A costs, unallowed expenses (e.g., subsidized employee meals), taxes, depreciation on facilities capital, payment delays, the fact that cost-plus contracts make progress payments on profit as well as on cost, ceilings on progress payments, as well as more complex forms of uncertainty (for example, about contract length as

---

[12]Note also that $\omega_w = 0$ if $r = 0$, confirming that there is no risk compensation for principal, only for interest.

Light: $\omega_w = \dfrac{1}{n(\mu - r)}[e^{\mu n} - e^{rn}] - \dfrac{1}{n\mu}[e^{\mu n} - 1]$ (4.17)

Heavy: $\omega_w = \dfrac{1}{2}rn$ (4.19)
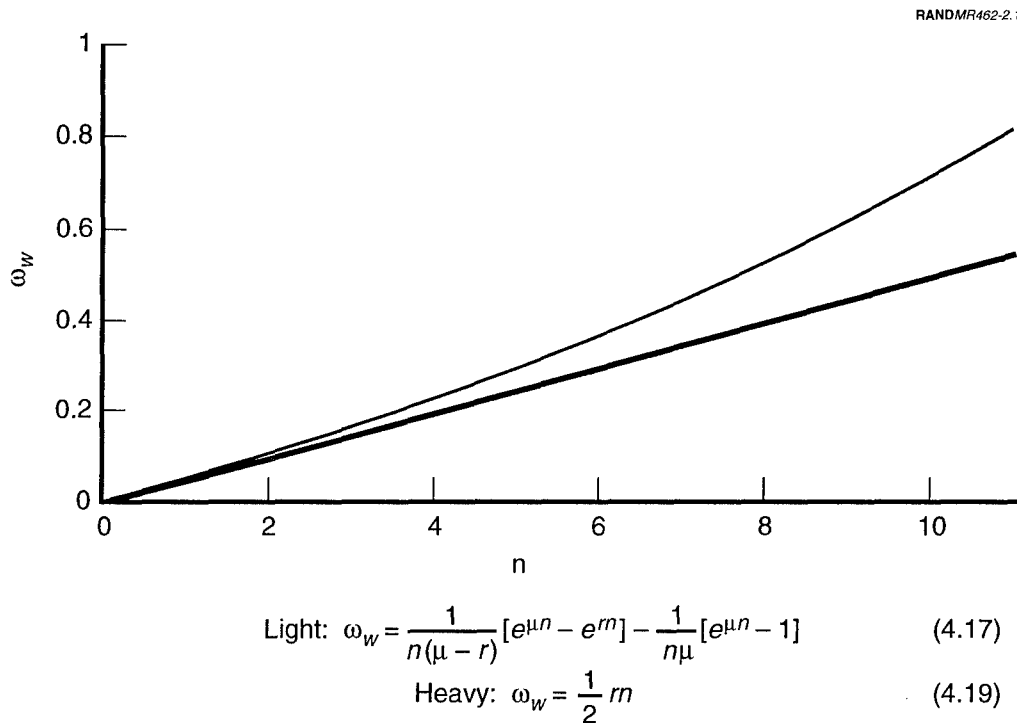
Figure 2.1—Corrections for Risk and Contract Parameters

well as about cost). Many of these concerns are addressed in Rogerson (1992a), Osband (1990a), or both. There are doubtless other complications to be considered as well. For example, Osband (1990a) argues that engineering or production teams have some of the qualities of facilities capital and therefore calculations similar to those for facilities might be applied to them.

In addition, the task remains of providing a theory for determining the optimal size of true profit, profit in excess of that needed to induce the firm to accept the contract. Something of a start on such a theory, but only a start, is represented by the calculation for the markup for facilities capital. Presumably, the overall true profit component may involve additional markups meant to induce investment in other areas, such as independent research. A complete theory of profit will require a better understanding of how profit policy interacts with the long-run decisionmaking of firms. Such interaction may lead to "counterintuitive" results, as in the observation that higher contract profit can mean lower overall profit from a long-run perspective; see the discussion of Bower and Osband (Chapter Four) in Section 2.

Measurement

The precision suggested by the derivations above is misleading. The formulas specify the form of the profit corrections needed, but estimation of the needed parameters is problematic. If for no other reason than the volatility of input prices, forecasting $\tilde{C}_t$ except over a short period is difficult. Estimating μ, the expected covariance of the

cost stream with the market portfolio, will be equally difficult. To a degree, μ, and $\tilde{C}_t$ can be estimated from the *ex post* performance of similar production contracts. However, the canonical profit policy contract is for the production of a singular, state of the art, piece of equipment and here, arguably, the value of history in making predictions may be limited.

Even computing the risk-free rate is not trivial. Although it can be proxied by the t-bill rate in the current period, for subsequent periods it must be inferred from the term structure of long-term rates. Rogerson (1992a) provides such a calculation involving the LIBOR futures market.[13] As Osband (1990b) notes, term structures, even for essentially riskless assets, appear to incorporate risk, apparently because of uncertainty over inflation. This may in turn mean that the data are inconsistent with application of the period-by-period CAPM; see Fama (1977). For discussion of other problems in estimating the "simple" static CAPM, all of which are inherited by the period-by-period CAPM, see Huang and Litzenberger (1988).

It perhaps should be stressed that these cautionary comments do not constitute an indictment of the derivations in the preceding subsections. The reader may note that the same estimation problems plague profit policy in any form it takes. The value of the derivations, and especially of Eq. (4.13), which arguably will hold under any "reasonable" model of asset pricing, is that they ensure that the basic framework is logically sound and provide some guidance as to exactly what needs to be estimated.

5. THEORETICAL VERSUS ACTUAL CONTRACT PROFIT

Following roughly the format of Section 4, the discussion will be broken into subsections covering working capital interest charges (π_w), facilities capital (π_{fac}), cost-sharing risk (π_c), and true profit and other compensation.

Working Capital

The minimum profit compensation for working capital is, from Eq. (4.12), of the form:

$$(1-\alpha)\omega_w C^e.$$

In continuous time we derived Eq. (4.17):

$$\omega_w = \frac{1}{n(\mu - r)} [e^{\mu n} - e^{rn}] - \frac{1}{n\mu} [e^{\mu n} - 1]. \tag{5.1}$$

A first-order approximation to this is Eq. (4.19):

---

[13]LIBOR is the London Interbank Offer Rate, the rate paid on three-month U.S. dollar time deposits in Europe.

$$\omega_w \approx \frac{1}{2} rn. \tag{5.2}$$

Thus, the risk component of working capital compensation is negligible for n small. As already noted in the previous section, the approximation in Eq. (5.1) accords fairly well with profit policy, which specifies roughly:[14]

$$\omega_w = \tau \left[\frac{1}{2} n - 0.35 \right] \tag{5.3}$$

provided that contract length is 76 months or less. Equations (5.2) and (5.3) differ in two respects. The first is that Eq. (5.3) uses the treasury rate τ, which runs a percentage point higher than the short-term riskless rate r. That is, if Eq. (5.1) and the theory behind it is taken seriously then the working capital adjustment may be too generous.[15] On the other hand, the ceiling on $L(n)$ and the subtraction of 0.35 work in the opposite direction. If $n = 3$, $r = 0.1$, and $\tau = 0.11$, we get $\omega_w = 0.150$ from Eq. (5.2) but $\omega_w = 0.127$ from Eq. (5.3), about 16 percent too low. As n grows, the term -0.35τ is dominated and the percentage error converges to $(\tau - r)/r$, which in this example is positive 10 percent. However, at these interest rates, the error is negative so long as the contract length n is less than 8 years. These errors, it should be borne in mind, are on top of the errors resident in the approximation itself. With $n = 3$, $r = 0.1$, and $\mu = 0.05$, Eq. (5.1) yields 0.175, so that Eq. (5.2) is 14 percent too low and Eq. (5.3) is 27 percent too low. See Figure 2.2.

Facilities Capital

From Eq. (4.10), if the probability that facilities will remain idle for one period is $1-\eta$, producers demand a minimum of:

$$\pi_{fac} = \left[(1 + r)^{n-1} \frac{(r + \eta)}{n} - 1 \right] K \tag{5.4}$$

in contract profit to invest \$$K$ in facilities capital. (Again, cost risk, and a number of other risks, were ignored in this analysis.) In continuous time:

$$\pi_{fac} = \frac{K}{\eta} \left[e^{rn} - (1 - \eta)e^{r(n-1)} - \eta \right]. \tag{5.5}$$

---

[14]As noted in Section 3, the weighted guidelines specify that on fixed-price contracts with no progress payments, $\alpha = 0$, firms should receive profit of 0.02 C^e. This is grossly wrong unless the contract is very short (a week).

[15]Note, however, that Eq. (4.16), hence Eq. (5.2), assumes that π_w is riskless. If π_w were risky, say because there were positive probability the government would refuse to accept delivery, then this risk would have to be compensated. Conceivably, the needed adjustment to the calculation might be approximated by discounting at a higher rate; see the discussion of discounting in Osband (1990a).

$$\text{Light: } \omega_w = \frac{1}{n(\mu - r)}[e^{\mu n} - e^{rn}] - \frac{1}{n\mu}[e^{\mu n} - 1] \qquad (5.1)$$

$$\text{Heavy: } \omega_w = \frac{1}{2}rn \qquad (5.2)$$

$$\text{Dotted: } \omega_w = \tau[\frac{1}{2}n - 0.35] \qquad (5.3)$$

Figure 2.2—Corrections for Risk and Contract Length

This is Eq. (C.3) from Appendix C, where two variants, arguably less plausible, are also derived. The first-order approximation to this about $n = 0$ is:

$$\pi_{fac} \approx \frac{K}{\eta}\left[(1 + nr)[1 - (1 - \eta)e^{-r}] - \eta\right]. \qquad (5.6)$$

Actual profit policy, on the other hand, is of the form:

$$\pi_{fac} = nrK + n\sum\omega_i K_i \qquad (5.7)$$

where i is specific to the type of capital (e.g., land, buildings, equipment). For land, $\omega_L = 0$.

Part of the difference between Eqs. (5.7) and (5.6) is illusory. If, for example, depreciation had been taken into account, then Eq. (5.4) would likewise depend on capital type. Conversely, the influence of η may be reflected in Eq. (5.7) through the terms ω_i. These complications aside, to reconcile Eqs. (5.6) and (5.7), one can derive that we must have, if there is a single capital good type:

$$\omega = \frac{1 - \eta}{\eta}(1 - e^{-r})\frac{(1 + nr)}{n}. \qquad (5.8)$$

Figure 2.3 plots ω as a function of η assuming $r = 0.1$ and $n = 3$. For $\eta = 1$ (the firm gets the contract for certain), $\omega = 0$ and

$$\pi_{fac} \approx nrK. \tag{5.9}$$

The appendix provides another story for why something like Eq. (5.9) might hold (see Eq. (C.6)). For $0 < \eta < 1$, ω is positive and increasing as η falls. Although, the weighted guidelines do not stipulate that ω be given by Eq. (5.8), DoD negotiators are given considerable latitude in specifying ω. Therefore, it is conceivable that, as implemented, profit policy is closer to the theory embodied in Eq. (5.6) than it may at first appear.

Cost-Sharing Risk

Recall that the profit compensation for cost-sharing risk is, from Eq. (4.10):

$$\gamma \omega_c C^e. \tag{5.10}$$

The weighted guidelines provide a risk adjustment, which is the sum of two terms. The "technical risk" term is of the form δC^e, which is not correct, since cost-sharing

RANDMR462-2.3

$$\omega = \frac{1 - \eta}{\eta} (1 - e^{-r}) \frac{(1 + nr)}{n} \tag{5.8}$$

Figure 2.3—Corrections for Risk in Employing Facilities Capital

risk must be a function of the level of cost sharing, γ.[16] In particular, technical risk gives profit to cost-plus-fixed-fee contracts (CPFF; $\gamma = 0$), which bear no cost-sharing risk. The "contract risk" term depends on contract type. The ranges for markups over cost have midpoints, which increase as one goes from CPFF contracts to firm-fixed-price (FFP) contracts (for which $\gamma = 1$). Thus, this component is roughly of the correct form.

The current regulations do not make ω_c explicitly a function of contract length n. Under the cost technology modeled here, taken from Osband (1990a), this is incorrect. The error introduced by ignoring contract length can be substantial. In the continuous time formulation of Eq. (4.16):

$$\omega_c = \frac{1}{n\mu}[e^{n\mu} - 1 - \mu n]$$

(5.11)

which by the first-order approximation of Eq. (4.18) is:

$$\omega_c \approx \frac{1}{2}\mu n.$$

(5.12)

That is, the risk adjustment should rise approximately linearly in contract length. Suppose that the risk adjustment coefficient is fixed at some ω_c^*. For specificity, suppose ω_c^* is chosen such that Eq. (5.12) is satisfied exactly at $n = 3$. Then, as an illustration, if $\mu = 0.05$:

$$\omega_c^* = 0.08.$$

However, if contract length were $n = 4$, then the value of ω_c given by Eq. (5.12) changes to:

$$\omega_c = 0.11.$$

Thus, even one additional contract year causes ω_c^* to underestimate ω_c by roughly 26 percent. This dwarfs, for example, the error from using Eq. (5.12). The omission of n from the profit policy specification of ω_c^* appears to be the main explanation for the finding in Osband (1989a) that the present value of profit policy contracts is sharply declining in n.

True Profit and Other Compensation

Few components in this category are explicitly represented in the weighted guidelines. As discussed in Section 3, certain additional expenses appear under management, but otherwise the corrections are loaded into the markups for cost risk, working capital, or facilities capital. This applies in particular to true profit, most of which

---

[16]But see the discussion of technical risk in Section 3.

must be computed as a residual.[17] Rogerson (1992a) argues that ideally, profit policy would provide explicit categories, and guidelines, for each of these items.

Using a somewhat different model of cost uncertainty than that used here, Rogerson (1992a) estimates that the current weighted guidelines provide true profit of 0.56 percent of cost on FFP contracts and 4.06 percent on CPFF contracts. It is not clear what the economic rationale would be for making true profit depend on the form of the contract. Rogerson suggests that the difference, a factor of 8, probably stems from the error, noted above, in giving compensation to CPFF contracts for cost-sharing risk even though they involve no cost sharing.

6. CONCLUSION

As the previous section detailed, the weighted guidelines appear to err in some recommendations for the "correct" level of compensation for unrecognized costs. The significance of this should not be exaggerated. The weighted guidelines, by permitting DoD negotiators to select cost markups from ranges that in many cases are quite broad, provide substantial room to address compensation problems. However, as Rogerson has emphasized, it may be desirable nevertheless to restructure profit policy so that it is more transparent in its operation and so that, in particular, what has been called the "true" profit component can more readily be determined. Improved transparency may lower the cost of contract negotiation and may in particular help forestall some of the lobbying that now takes place.

The theory of the weighted guidelines sketched here is not complete but it is considerably advanced. Research on profit policy may now be most productive if directed at neighboring issues. Of these, two seem especially worthy of attention. First, there has been relatively little research on how joint costs, for example, buildings shared by more than one project, are allocated across contracts. One conjecture is that, because negotiated profit is computed as a percentage of cost, firms may bargain with DoD negotiators to have the charges for joint costs assigned disproportionately to profit policy contracts. It remains for future work to determine whether this actually happens, and what the net efficiency consequences are. Second, more work remains to be done on profit policy from the perspective of long-term industry decisionmaking. Broad issues include DoD commitment to stable policy and the overall role of negotiated contracts in procurement. More narrowly, if "true" contract profit is to be used as a policy instrument, then its effects must be gauged and a method found for setting it at an appropriate level.

APPENDIX A: CONTRACT TYPES

There are two basic contract types, "fixed price" and "cost plus." These are defined in reference to extreme representatives: the FFP contract and the CPFF contract.

---

[17]The component labeled cost control is a reward for perceived managerial excellence on previous contracts, and so may, in a sense, be considered part of true profit. But see the discussion of cost control in Section 3.

The CPFF contract sets the cost-sharing parameter $\gamma = 0$ (full cost sharing) and the progress payment rate $\alpha = 1$. Moreover, progress payments apply to both cost reimbursement and contract profit. Consequently, with payment being made in full during the course of the contract, the payment p at the end of the contract is 0.

In contrast, under the FFP contract, $\gamma = 1$ (zero cost sharing) and $\alpha \in [0,1]$, so that Eq. (3.1) becomes:

$$p = (1-\alpha)C + \pi^* - C + C^e$$
$$= \pi^* + C^e - \alpha C.$$

In words, the payment at contract end is contract profit π^* plus expected cost less the fraction of actual cost paid out as progress payment. Intermediate between these two extremes are CPIF and FPI contracts.

Like the CPFF, a CPIF sets $\alpha = 1$ and progress payments apply to both cost and profit. Unlike the CPFF, the sharing ratio γ is set above zero; 15 percent is a typical sharing rule. Under a cost plus incentive fee contract, γ reverts to zero if actual costs fall outside of some interval around C^e. Thus, in particular, the firm bears γ of cost overruns up to some upper bound $C^e + b$, but bears 0 of the cost overrun beyond that.

The fixed-price-incentive contract sets $\alpha \in [0,1]$, $\alpha = 80$ percent is typical, and sets cost share at a value less than 1, but usually greater than the cost share on a CPIF contract; a γ of 30 percent on a FPI contract is typical. As noted in the text, progress payments on a FPI contract apply only to cost, not to profit. Finally, if realized costs fall outside of some interval, FPI contracts revert to firm-fixed-price contracts (whereas CPIF contracts revert to CPFF). Thus, the firm bears γ of cost overruns up to some $C^e + b$, where b is some positive number, and all of the cost overrun beyond that. Unlike with CPIF contracts, underruns do not cause a similar reversion.

APPENDIX B: UNCERTAINTY

Consider first the static CAPM of introductory finance in which there are only two periods: Assets are traded today (date 0) and dividends are paid out tomorrow (date 1); see Huang and Litzenberger (1988). Consider some asset, say a stock, and denote its return over the period by $\tilde{\rho}$ and expected return by ρ^e. Let the return on the market portfolio be $\tilde{\rho}^e_M$ and its expectation ρ^e_M. "Market portfolio" is defined in the text. Recall that we have assumed that there is a constant riskless rate of return of r per period, where riskless means that r is not random.

Under the assumptions of the capital asset model, equilibrium requires that:

$$\rho^e - r = \frac{\text{cov}(\tilde{\rho}, \tilde{\rho}_M)}{\text{var}(\tilde{\rho}_M)} \left(\rho^e_M - r \right) \tag{B.1}$$

where cov denotes covariance and var denotes variance. In words, the risk premium $(\rho^e - r)$ on the asset is exactly proportional to the risk premium on the market as a

whole, $(\rho_M^e - r)$, with the factor of proportionality being the normalized covariance of the asset return with the market return $\text{cov}(\tilde{\rho}, \tilde{\rho}_M)/\text{var}(\tilde{\rho}_M)$. Thus, CAPM says that the market requires a premium for randomness (as reflected in a higher ρ^e) only to the extent that the asset is correlated with the market as a whole. The variance of $\tilde{\rho}$ per se does not matter. Part of the intuition for this is that what investors care about is not the variability of an asset in isolation, but how that variability affects the total variability of their asset portfolios. For a more deliberate and careful exposition, the reader should consult Huang and Litzenberger (1988).

For our purposes, it is convenient to rewrite Eq. (B.1) so that it is in terms of the current price of the asset rather than the asset's rate of return. Let V be the price of the asset and $\tilde{X}$ be its gross return, so that $\tilde{\rho} = (\tilde{X} - V)/V$. Substituting into Eq. (B.1) yields:

$$\left(\frac{\tilde{X}^e}{V} - 1 - r \right) = \frac{\text{cov}\left(\frac{\tilde{X}}{V} - 1, \tilde{\rho}_M \right)}{\text{var}\left(\tilde{\rho}_M \right)} \left(\rho_M^e - r \right)$$

or,

$$\left(\frac{\tilde{X}^e}{V} - (1+r) \right) = \frac{1}{V} \frac{\text{cov}\left(\tilde{X}, \tilde{\rho}_M \right)}{\text{var}\left(\tilde{\rho}_M \right)} \left(\rho_M^e - r \right). \tag{B.2}$$

Let:

$$\lambda = \frac{\left(\rho_M^e - r \right)}{\text{var}\left(\tilde{\rho}_M \right)}.$$

Then Eq. (B.2) can be manipulated to yield:

$$V = \frac{X^e - \lambda \text{cov}\left(\tilde{X}, \tilde{\rho}_M \right)}{\left(1 + r \right)}$$

$$= X^e \left[\frac{1 - \lambda \text{cov}\left(\frac{\tilde{X}}{X_e}, \tilde{\rho}_M \right)}{\left(1 + r \right)} \right]. \tag{B.3}$$

In words, the present value of the asset paying $\tilde{X}$ is the discounted expected value of X^e with a "risk-adjusted" discount rate given by:

$$1 + r^* = \frac{1+r}{1 - \lambda \operatorname{cov}\left(\dfrac{\tilde{X}}{X^e}, \tilde{\rho}_M\right)}. \tag{B.4}$$

Although the discussion to this point has been in terms of the return on stock, the same algebraic machinery applies to any random asset, including a random cost. If the cost at date 1 is $-C$, this is worth at date 0:

$$V = -C^e \left[\frac{1 - \lambda \operatorname{cov}\left(\dfrac{\tilde{C}}{C^e}, \tilde{\rho}_M\right)}{1+r}\right]. \tag{B.5}$$

As in the text, the assumption will be made that the sign of $\operatorname{cov}(\tilde{C}, \tilde{\rho}_M)$ is negative: Higher defense industry costs tend to be correlated with lower profits in the economy as a whole. Thus, confirming the obvious, the present value of the date 1 cost $\tilde{C}$ is unambiguously negative.

We want to derive the analog to Eq. (B.5) when the cost is incurred in period $t \geq 1$. To this end, we wish to use Eq. (B.5) recursively. At date $t - 1$, the value of the cost incurred in date t is:

$$V_{t-1}^t = -E_{t-1}(\tilde{C}_t) \left[\frac{1 - \lambda_{t-1} \operatorname{cov}_{t-1}\left(\dfrac{\tilde{C}_t}{E_{t-1}(C_t)}, \tilde{\rho}_{M_t}\right)}{1+r}\right] \tag{B.6}$$

where V_{t-1}^t is read, "The value at date $t - 1$ of a cost incurred at date t." E_{t-1} denotes expectation based on information known at date $t - 1$, cov_{t-1} denotes covariance based on information known at date $t - 1$, and $\lambda_{t-1} = (E_{t-1}(\tilde{\rho}_{M_t}) - r)/\operatorname{var}_{t-1}(\tilde{\rho}_{M_t})$. Applying Eq. (B.5) again:

$$V_{t-2}^t = -E_{t-2}(\tilde{V}_{t-1}^t) \left[\frac{1 - \lambda_{t-2} \operatorname{cov}_{t-2}\left(\dfrac{\tilde{V}_{t-1}^t}{E_{t-2}(\tilde{V}_{t-1}^t)}, \tilde{\rho}_{M_t}\right)}{1+r}\right]. \tag{B.7}$$

In principle, $E_{t-1}(\tilde{C}_t)$, $\operatorname{cov}_{t-1}(\cdot, \cdot)$, and λ_{t-1} could all be date $t - 2$ random variables. Thus, substitution of Eq. (B.6) into Eq. (B.7) could yield an extremely complex expression. However, internal consistency of the period-by-period CAPM imposes strong restrictions on what can in fact be random at $t - 2$. Following Fama (1977), we assume that only $E_{t-1}(\tilde{C}_t)$ is random. For simplicity, we further assume that λ_t is a

constant. Then, upon substituting Eq. (B.6) into Eq. (B.7), we derive, since $E_{t-1}[E_{t-1}(\tilde{C}_t)] = E_{t-2}(\tilde{C}_t)$:

$$V_{t-2}^t = -E_{t-2}(\tilde{C}_t)\frac{1}{(1+r)^2}\left[1-\lambda\,\mathrm{cov}_{t-2}\left(\frac{E_{t-1}(\tilde{C}_t)}{E_{t-2}(\tilde{C}_t)},\tilde{\rho}_{M_{t-1}}\right)\right]$$
$$\left[1-\lambda\,\mathrm{cov}_{t-1}\left(\frac{\tilde{C}_t}{E_{t-1}(\tilde{C}_t)},\tilde{\rho}_{M_t}\right)\right].$$

(B.8)

Continuing in this manner:

$$V_0^t = -E_0(\tilde{C}_t)\frac{1}{(1+r)^n}\prod_{i=0}^{t-1}\left[1-\lambda\,\mathrm{cov}_i\left(\frac{E_{i+1}(\tilde{C}_t)}{E_i(\tilde{C}_t)},\tilde{\rho}_{M_{i+1}}\right)\right].$$

(B.9)

Thus, the period-by-period CAPM discounts at the riskless rate a risk-adjusted expectation of $\tilde{C}_t$.

Assume that $\tilde{C}_t$ can be written as the sum of some period 0 cost plus a sequence of cost shocks, say to input prices or to the production technology. Following Osband (1990a), if $\tilde{C}_t$ is independent of the order in which these shocks arrive, then:

$$\mu_{it} = -\lambda\,\mathrm{cov}_i\left(\frac{E_{i+1}(\tilde{C}_t)}{E_i(\tilde{C}_t)},\tilde{\rho}_{M_{i+1}}\right)$$

will be a constant, independent of i. Assume further that it is independent of t. Note that since we assume, $\mathrm{cov}(\tilde{C},\tilde{\rho}_M) < 0$, we have $\mu_{it} = \mu > 0$. Substituting into Eq. (B.9):

$$V_0^t = -E_0(\tilde{C}_t)\left[\frac{1+\mu}{1+r}\right]^r.$$

(B.10)

Thus the present value of the total cost stream is:

$$V = \sum_{t=1}^n V_0^t = -(1-\alpha)\frac{C^e}{n}\sum_{t=1}^n\left[\frac{1+\mu}{1+r}\right]^t.$$

(B.11)

To price out the payment, note first that we can write:

$$p = (1-\alpha)\tilde{C} + \pi_m - \gamma(\tilde{C} - C^e)$$
$$= \pi_m + \gamma C^e + (1-\alpha-\gamma)\tilde{C}.$$

The term $\pi_m + \gamma C^e$ is assumed nonrandom, hence does not covary with $\tilde{\rho}_{M_t}$, hence is discounted at the riskless rate r. The term $(1-\alpha-\gamma)\tilde{C}$ is the sum of terms $(1-\alpha-\gamma)\tilde{C}_t^n$, where $\tilde{C}_t^n$ is $\tilde{C}_t$ at date n. From the perspective of any period $s \geq t$, $\tilde{C}_t^s$

is known, hence gets discounted at the riskless rate. Beginning at date $t-1$, however, $\tilde{C}_t$ is random and is evaluated as in Eq. (B.10). Thus, the date 0 value of $\tilde{C}_t^n$ is

$$\frac{1}{(1+r)^{n-t}} E_0(\tilde{C}_t) \left[\frac{1+\mu}{1+r}\right]^t.$$

Combining, the date 0 value of the payment p is:

$$W = \frac{\pi_m + \gamma C^e}{(1+r)^n} + \frac{(1-\alpha-\gamma)}{(1+r)^n} \frac{C^e}{n} \sum_{t=1}^{n}(1+\mu)^t$$

$$= \frac{1}{(1+r)^n} \left\{ \pi_m + (1-\alpha)\frac{C^e}{n}\sum_{t=1}^{n}(1+\mu)^t - \gamma\frac{C^e}{n}\left[\sum_{t=1}^{n}(1+\mu)^t - n\right]\right\}.$$

Setting $V + W = 0$ yields:

$$\frac{1}{(1+r)^n}\left\{\pi_m + (1-\alpha)\frac{C^e}{n}\sum_{t=1}^{n}(1+\mu)^t - \gamma\frac{C^e}{n}\left[\sum_{t=1}^{n}(1+\mu)^t - n\right]\right\}$$

$$= (1-\alpha)\frac{C^e}{n}\sum_{t=1}^{n}\left[\frac{1+\mu}{1+r}\right]^t$$

or

$$\pi_m = \gamma\frac{C^e}{n}\left[\sum_{t=1}^{n}(1+\mu)^t - n\right] + (1-\alpha)\frac{C^e}{n}\left\{\sum_{t=1}^{n}\left[\frac{1+\mu}{1+r}\right]^t - \sum_{t=1}^{n}(1+\mu)^t\right\}. \qquad (B.12)$$

From this, we get immediately Eqs. (4.14) and (4.15).

APPENDIX C: FACILITIES PROFIT IN CONTINUOUS TIME

Suppose that interest compounds continuously but otherwise there is no basic change in the model. In particular, it remains the case that if at time 0 the firm fails to win a contract, it must wait one full period (one year) either to sell its facilities or bid for a new contract. Given this, if we divide each period into s subperiods, Eq. (4.9) becomes:

$$K = \eta(\pi_{fac} + K)\frac{1}{(1+r/s)^{sn}} + (1-\eta)\frac{K}{(1+r/s)^s} \qquad (C.1)$$

which can be rewritten as:

$$\pi_{fac} = \frac{K}{\eta}\left[(1+r/s)^{sn} - (1-\eta)(1+r/s)^{s(n-1)} - \eta\right].$$ (C.2)

As *s* goes to infinity, this approaches:

$$\pi_{fac} = \frac{K}{\eta}\left[e^{rn} - (1-\eta)e^{r(n-1)} - \eta\right].$$ (C.3)

A first-order approximation to this around $n = 0$ is given by:

$$\pi_{fac} \approx \frac{K}{\eta}\left[(1+nr)[1-(1-\eta)e^{-r}] - \eta\right].$$ (C.4)

An alternative is to model the period itself as shrinking. In this case, failure to win a contract at date 0 means that the firm must wait only until date $1/s$ to bid on a new project or sell its capital. Thus, as *s* goes to infinity, the length of time that a firm's facilities capital investment remains "sunk" shrinks to zero. This is contrary to the spirit of the model sketched in Section 3, but the approach yields somewhat simpler formulas. Moreover, one variant (Eq. (C.7) below), is the approach actually employed in Osband (1990a).

Suppose that the probability of winning a contract holds fixed at η. Then Eq. (C.1) becomes:

$$K = \eta(\pi_{fac} + K)\frac{1}{(1+r/s)^{sn}} + (1-\eta)\frac{K}{1+r/s},$$ (C.5)

which can be rewritten as:

$$\pi_{fac} = \frac{K}{\eta}\left[(1+r/s)^{sn} - (1-\eta)(1+r/s)^{sn-1} - \eta\right].$$

As *s* goes to infinity, this approaches:

$$\pi_{fac} = K\left[e^{rn} - 1\right].$$

Note that this does not depend on η, the probability of winning a contract in a given period. As *s* goes to infinity the firm is allowed to bid "infinitely often" in any given time interval. Because the probability of winning a contract in any given auction is fixed at $\eta > 0$, the firm will almost surely get a contract "infinitely close" to date 0. A first-order approximation to this around $n = 0$ is given by:

$$\pi_{fac} \approx rnK.$$ (C.6)

It may be noted that under this scenario, the continuous time analog of Eq. (4.11) gives a π' equal to π_{fac}: There is no room for the government to exploit firms that have already "sunk" facilities investments. The reason is simply that the time period is now instantaneous and the probability of winning a contract is effectively 1.

On the other hand, we could view the probability of winning a contract at any date as η/s. Thus the probability of winning a contract in a given period shrinks proportionately with the length of the period, much as interest per period shrinks (to r/s). The idea is that as the period length shrinks, contracts become available more frequently but the number of competitors increases as well, so that the probability of winning a contract in a given period grows smaller. Equation (C.5) becomes:

$$K = (\eta/s)(\pi_{fac} + K)\frac{1}{(1+r/s)^{sn}} + (1-\eta/s)\frac{K}{1+r/s}, \tag{C.7}$$

which can be rewritten as:

$$\pi_{fac} = K\left[(1+r/s)^{ns-1}\frac{r+\eta}{\eta} - 1\right].$$

As s goes to infinity, this converges to:

$$\pi_{fac} = K\left[e^{rn}\frac{r+\eta}{\eta} - 1\right].$$

As noted, this is the continuous time formulation actually used in Osband (1990a). A first-order approximation to this around $n = 0$ is given by:

$$\pi_{fac} \approx \frac{K}{\eta} r \left[n(r+\eta)+1\right]. \tag{C.8}$$

PROFIT REGULATION OF DEFENSE CONTRACTORS AND PRIZES FOR INNOVATION
by William P. Rogerson[1]

1. INTRODUCTION

This paper first argues, based on theoretical grounds, that informational and incentive constraints inherent in the innovation process require that regulatory institutions in defense procurement create prizes for innovation. Since the quality of an innovation is difficult to objectively describe or measure, the most natural method for awarding prizes is to allow firms to earn positive economic profit on production contracts. Explicit recognition of this role of profit regulation generates interesting perspectives on a number of important policy issues involving regulatory design. The value of the prizes offered on twelve major aerospace systems are calculated. The prizes are clearly large enough to support the contention that their existence is an important aspect of the current regulatory structure.[2]

Section 2 of the paper presents background information on the procurement process. Section 3 then presents a more complete description of the theory. Section 4 shows that this simple idea sheds new light on a number of current policy debates on how profit policy should be structured. Section 5 develops the theoretical basis for using observed changes in firms' stock market value to infer the size of the prizes they were competing for. Section 6 presents the empirical measurement of the changes in firms' stock market values. Section 7 uses the theory of Section 5 and the data of Section 6 to estimate the size of the prizes.

2. BACKGROUND

In a typical aerospace project, the DoD sponsors a design competition where two or more firms are funded to independently produce competing proposals. Depending

---

[1]This paper is reprinted from William P. Rogerson, "Profit Regulation of Defense Contractors and Prizes for Innovation," the *Journal of Political Economy*, Vol. 97, No. 6, December 1989, pp. 1284–1305. Copyright © 1989 by The University of Chicago. Used by permission. A longer and more detailed version of this paper is available as a RAND report, R-3635-PA&E, *Profit Regulation of Defense Contractors and Prizes for Innovation*, 1992.

[2]In a related paper, Lichtenberg (1988a) has directly attempted to estimate the amount of privately funded R&D induced by military procurement.

on the program, competing firms are sometimes asked to submit actual functioning prototypes or sometimes are simply required to submit detailed studies. The winning firm goes on to build or adapt a production line[3] and becomes the sole prime contractor for the program. Economies of scale together with very small production runs render it economically infeasible to have two or more firms build fully functioning production lines.

As part of their design proposals, firms will typically be required to bid on the first one to three years of production. However, all subsequent purchases are negotiated by the firm and the DoD on an annual basis. In fact, if the winner is chosen at a very early stage of the process based perhaps only on paper design proposals, the prices for all production runs may be left to be determined by subsequent negotiations. Transactions costs together with constantly evolving technologies and requirements are thought to render long-term contracts infeasible.

A complex set of regulations, often referred to as profit policy, describe how the government negotiator should calculate a "fair price" in such noncompetitive procurements. Pricing is cost-based, as in public utility regulation, although the regulations used in defense procurement differ in a number of respects from the rules used in public utility regulation.[4] The regulations can be viewed as establishing an administered contract between the DoD and defense firms.[5] The administered contract is that when individual contracts are negotiated, the DoD's negotiation objective will *not* be to get the absolutely lowest price possible on that individual contract. Rather, the DoD's objective will be to pay the contractor a "fair" price where "fair" is determined by existing profit policy regulations.

This paper investigates whether economic profit is earned by the prime contractor on work that occurs after the winner of the design contest is announced. This includes work performed under contracts that are signed the day the winner is announced as well as all the contracts that will be subsequently negotiated.

3. THEORY

The standard regulated public utility is engaged in a one-step process—production. The regulator attempts (among other things) to guarantee that the regulated firm earns zero economic profit on this step. The defense sector differs fundamentally from this standard paradigm because it is engaged in a two-step process—innovation and production. The first step is at least as important as the second. In fact, a basic assumption underlying current U.S. defense policy is that the optimal strategy involves having a smaller number of more technically sophisticated weapons than the enemy. Given this strategy, it is critically important for the United States to maintain an innovative lead. Thus, the DoD must find some way to induce defense contractors to exert large amounts of effort directed toward generating the types of innova-

---

[3]This pre-production activity is called full-scale development.

[4]See Rogerson (1992a) for a detailed analysis of these regulations.

[5]See Goldberg (1976) for a full discussion of the idea of regulation as an administered contract.

tions that DoD would find most useful. A good regulatory policy would still presumably attempt to guarantee that regulated defense firms earn zero economic profit overall. However, *a priori* there is no reason to require that firms earn zero profits on each step. In fact, the major theoretical point of this paper is that there is a very good reason to structure the regulatory process so that negative economic profit is earned in the innovation phase and positive profit is earned in the production phase.

The theory is most easily explained by segmenting it into four parts.

Part 1: Prizes for innovation are required.

The argument of Part 1 is that the DoD is unable to directly purchase the innovative efforts of firms. Therefore, it must indirectly give firms the incentive to provide this effort by establishing rewards for successful innovation. This is true for two reasons.

First, there is a moral hazard problem. Since it is difficult to monitor the level of effort a firm is exerting, the DoD must give firms an incentive to exert this effort by promising to reward successful innovation with prizes.

However, even if level of effort was totally observable, a second factor would still necessitate the use of prizes. This is that firms are very likely to possess private information about which sorts of projects are more likely to yield the kind of results of most value to the DoD. To illustrate this idea, suppose that exerting effort consisted only of spending money and the DoD could exactly monitor the amount of money spent. Furthermore, assume that two possible projects exist for a firm to explore—projects A and B. Assume as well that the DoD can also monitor whether money is spent on project A or B. Therefore, the DoD could simply directly order the firm to exert given levels of effort on each project and directly monitor that this occurred. However, now suppose as well that project A is likely to produce high benefits to the DoD but will yield very few commercial spinoffs for the firm. Project B is the reverse. It is likely to produce very low benefits for the DoD but will yield a number of useful ideas that the firm can use in its commercial business. Furthermore, suppose that because of its greater technical expertise, only the firm is aware of this fact. Both projects appear to be similar to the DoD. If the firm were simply hired to perform research (which is possible by assumption because research effort is directly monitorable) the firm would have an incentive to recommend project B. To give the firm an incentive to choose project A, the DoD must pay the firm not according to the amount of effort it exerts but instead according to the value of the ideas produced. That is, successful innovations must be rewarded with some sort of prize.

Another way of stating this second point is that an optimal research program should be somewhat decentralized so that firms can make decisions based upon their private information. However, when delegating some decisionmaking to firms, the DoD must simultaneously provide firms with incentives to make the decisions that are best from the DoD perspective. Establishing prizes for innovation accomplishes this. An example of this is that companies will often fund prototypes or at least initial research efforts for a particular system that they believe has great potential even if no

one in DoD at that time yet agrees.[6]   Thus, when there are prizes for successful innovation, firms have an incentive to use their own funds if necessary to pursue research projects that they strongly believe will eventually yield results of great value to the DoD.

Part 2: A regulatory structure that directly provides larger prizes for higher-quality innovations is not possible.

The argument of Part 1 does not by itself establish the regulatory principle that defense firms should earn positive profits on production contracts. In principle, government could commit to R&D incentive contracts of the form $w(x)$ where $x \in X$, X is the space of all possible innovations, and $w(x)$ is the wage the contractor will receive if innovation x results. Then, production contracts could be priced to yield zero economic profit and the payment of higher wages to more valuable innovations would provide the incentive for innovation. Furthermore, $w(x)$ could be chosen so firms were just willing to perform the R&D and thus earned zero economic profit in the R&D phase as well.

However, it is clear that the transactions costs of writing out a legally enforceable objectively verifiable contract describing all possible innovations and the price that would be paid for each one would be prohibitively costly if not impossible for all but the most trivially simple R&D projects. Some R&D occurs within well-defined programs with fairly well-defined objectives. Even in these cases, it seems unlikely that DoD could provide a legally enforceable contract covering all possible design improvements. However, a large fraction of firms' R&D is directed toward identifying more basic new ideas and concepts for weapons development. As explained in Part 1, the R&D process is somewhat decentralized to allow firms to use their own private information in deciding which avenues of R&D to explore. To sign a legally enforceable contract directly rewarding the results of this more far-ranging basic R&D would literally require government to list the possible universe of innovations and the prize attached to each one. This is obviously impossible.

One other option would be for government to simply announce that it would evaluate the quality of each new innovation and award a prize based on the evaluated quality. One might imagine creation of a "DoD prize panel," which annually assessed the results of all firms' efforts and awarded prizes accordingly. Such a scheme would probably be totally infeasible because of the subjectivity of any such evaluation. Firms might all claim that their research was unfairly evaluated and one could imagine endless congressional investigations into such a scheme. (It might also be politically difficult to award large prizes.)

Part 3: Contracts that provide economic profit on production contracts will provide prizes that are correlated with the importance of the innovation.

The obvious objectively verifiable signal of whether a firm has created a successful new weapon design is whether the DoD chooses to purchase it. Thus, a regulatory

---

[6]For example, a number of prototype predecessors of the F-16 and AH-64 were privately funded by firms. See Smith et al. (1981), pp. 84 and 155–158.

system could create prizes for innovation by guaranteeing that any firm that becomes a prime contractor on a new weapon system will earn positive economic profit on the production contracts for the weapon system. In such a system, firms that can successfully generate ideas good enough to be adopted by the government would receive prizes in the form of economic profit on the production phase of the system.

Furthermore, if profit is awarded approximately as a percentage of cost (i.e., the profit earned on a system doubles if the system is twice as expensive), this might in a very rough sense also tend to award larger prizes for better innovations for two reasons. First, systems that prove to be useful will be purchased in larger quantities. Second, there is probably some sense in which a $30 billion project is more important to government than a $30 million project.

Finally, note that it is important that any regulatory system provide the firm with incentives to devote effort to innovation even after it is selected as the prime contractor. This is for two reasons. First, since the prime is typically chosen before full-scale development, it must often perform significant development work before commencement of production. Second, constant upgrading of systems that are in production is a very important part of the overall innovative effort in defense procurement. A regulatory system that provides economic profit approximately equal to a percentage of cost provides incentives for both types of effort. During full-scale development, the lure of positive economic profit on production contracts will provide the firm with an incentive to exert effort toward ensuring the success of the project. During production, a firm that succeeds in improving the system will guarantee more sales and thus more profit.

Thus, the existence of an explicit design competition between two or more firms is not essential for economic profit to create incentives for innovation. What is essential is that the firm perceive that its chances of being awarded production contracts will increase if it exerts more innovative effort. An explicit design competition in which only one winner will be chosen clearly does this. However, even after the prime is chosen for a system, the program continues to compete for funding with other programs and thus the incentives to innovate still persist, though possibly at a reduced level.

Part 4: A rent-seeking model describes the equilibrium response of innovation and overall profit to prize levels.

In a rent-seeking model of the sort originally analyzed by Tullock (1967), firms spend money attempting to win a prize. Even over a time horizon where entry and exit are impossible, increasing the prize level by one dollar will not simply cause firms' profits to increase by one dollar. Rather, some fraction of the dollar will be channeled into increased expenditures devoted to attempting to win the prize. Over a time horizon where entry and exit are possible, changing the prize level has no effect at all on profits. When the prize level is increased by one dollar, existing firms will have an incentive to spend more money attempting to win the now larger prize. If the increase in their expenditures is less than one dollar, then they will be earning positive expected profits and more firms will enter. The new entrants will also spend money

attempting to win the prize. Entry will occur until aggregate rent-seeking expenditures equal the size of the prize and firms are earning zero expected profits overall.[7]

This rent-seeking formulation clearly applies to the regulatory structure in defense procurement. The existence of prizes or economic profit in the production phase of weapon programs induces firms to spend their own money attempting to win the right to produce weapon systems. Innovation and congressional and executive branch lobbying are probably the two primary forms of rent-seeking activities available to defense firms.

4. IMPLICATIONS FOR REGULATORY POLICY

The purpose of this section is to outline a number of implications that this theory has for how procurement policy should be optimally structured. It does not attempt to completely and formally explore all the implications of this theory for regulatory policy. Such an analysis would require a detailed description and consideration of the current rules for determining prices, which is beyond the scope of this paper. Rather, it simply attempts to show that the idea of "prizes for innovation" is an extremely useful organizing principle when thinking about regulatory issues in defense contracting.

Implication 1:   An important function of profit policy is to control the rate of innovation.

An ongoing and seemingly never-ending debate in policymaking circles in Washington concerns whether defense profits are too high or too low and whether price levels allowed under profit policy should therefore be adjusted to correct this problem. Every few years Congress (usually through the GAO) or the DoD produces a new calculation of accounting rates of return and the debate begins anew.[8] The debate over whether price levels should be raised or lowered then focuses on the problems with using accounting-based numbers and over what a "fair" profit rate should be. However, the implicit assumption, which all sides in the debate seem to agree with, is that the only effect of raising (lowering) allowed profit levels on production by some amount would be to raise (lower) firms' overall profit levels by the same amount and possibly induce some entry or exit. Since profit policy is seen simply as a tool for regulating firms' overall profit levels, the entire debate thus focuses on whether overall profits appear too high or too low.

If this paper's theory is correct, an important function of profit policy may be to regulate the level of innovative activity in the defense sector. Therefore, an important focus of the debate should be whether an adequate level of innovation currently exists or not. Even if entry and exit into the defense sector were impossible, a share of any increase in profit levels earned on production contracts will be transformed into increased innovative activity. If entry and exit are possible, long-run profits will necessarily be zero. Thus, profit policy has no effect at all on long-run profits. Rather, the

---

[7]See Rogerson (1992b) for a formal rent-seeking model that illustrates these points.

[8]The most recent studies are Comptroller General of the United States (1986) and DoD (1985).

only long-run effect of profit policy is to determine the level of rent-seeking expenditures.[9]

Implication 2: The correct regulatory principle may be to set price equal to full cost plus a percentage of full cost where "full cost" means operating cost plus capital cost.

The first feature of such a rule is that the prize is bigger for projects that involve greater expense. This might be thought to roughly insure that firms producing more useful innovations will receive larger prizes for the reasons discussed in Part 3 of Section 3. The second desirable feature is that the firm has no incentive to distort its mix of capital and noncapital expenditures away from the minimum cost ratio. This is because the rule provides an equal profit for the firm regardless of whether the firm spends it on operating costs or capital costs.[10]

Implication 3: There may be a tradeoff between encouraging innovation and encouraging productive efficiency.

The proposed rule in Implication 2 has the property that firms incurring higher costs will also earn higher profits. This clearly gives firms the incentive to attempt to maximize production cost once they have been selected to produce a system. Thus, the most natural method of implementing a prize system may also, unfortunately, create disincentives for firms to minimize production cost. The question of how to design pricing rules that simultaneously attempt to deal with the problems of creating prizes for innovation and incentives to minimize production cost is clearly an important topic for future research.

Implication 4: Policies such as dual sourcing, which reduce economic profit on production contracts, may reduce innovation.[11]

Implication 5: The current regulatory system encourages vertical integration of innovation and production.

It is clear that firms that design new weapons will have an incentive to integrate downstream into production if the rewards for excellent designs are in the form of profits on production contracts. Thus, the vertical integration of the R&D and production functions in the United States defense industry may be due to the regulatory structure rather than to any natural economic advantage of performing both functions within the same firm. Alexander (1973) suggests, for example, that Lockheed's research and production are quite separately organized. He also states that design and production are separately organized in the former USSR and that large prizes are directly paid to successful designers.[12]

---

[9]Of course, if entry lags are large, it may be that profit policy has fairly long-lasting effects on profit levels as well as innovation and that both effects need to be considered. The point of this paper is that the role of profit in creating incentives for innovation should certainly be one of the factors explicitly considered when evaluating possible changes to profit policy.

[10]See Bailey (1973).

[11]See Anton and Yao (1990), and Riordan and Sappington (1989) for theoretical models of dual sourcing.

[12]See pp. 9–10.

Implication 6: Different pricing rules may be appropriate for different sectors of the defense industry.

Current profit policy rules are intended to apply uniformly to all defense contracts. However, the need and importance of innovation clearly varies among sectors in the defense industry. This suggests that pricing rules should vary from sector to sector depending on how much innovative activity is required. In particular, pricing rules for more standard products should provide less economic profit.

Implication 7: It could be difficult to provide adequate prizes if the defense sector was publicly owned.

It is periodically suggested that the defense sector should be nationalized.[13] The theory of this paper suggests a possible problem with this idea. Namely, it is probably difficult to award large prizes to executives of nationalized companies.

Implication 8: The role of the IR&D program.

The DoD encourages firms to spend money on innovation not only by creating prizes but also by directly subsidizing this activity. Through the IR&D program, the DoD pays for an agreed-upon fraction of defense firms' expenditures on independently chosen and conducted research programs.[14] The theory of this paper suggests two possible reasons why this extra policy may be useful.

First, as explained in Section 3, a problem with the policy of inducing innovation through prizes is that it tends to cause other less-desirable forms of rent-seeking behavior in addition to innovation. Thus, it would be desirable to attempt to create a policy that channelled more of the rent-seeking behavior toward innovation. The obvious method for doing this is to subsidize expenditures on innovation and tax expenditures on other rent-seeking activities such as lobbying. The IR&D program accomplishes the former objective. The latter objective is accomplished to some extent by regulations that make lobbying expenditures unallowable for purposes of costing defense contracts.[15]

Second, compared to private industries where innovation is important, such as the pharmaceutical industry, the prize levels in defense probably do not vary as much with the quality of the innovation. A firm producing an extremely useful new drug may earn profit margins of well over 100 percent. However, for reasons outlined in Section 3, defense firms earn relatively constant profit margins regardless of the value of their innovation. This suggests that it may be the case that defense firms do not have adequate incentives to devote sufficient resources to researching the absolutely most important programs to the DoD. Subsidizing research in these most important areas may substitute for extremely large prizes.

---

[13]See, for example, Galbraith (1969).

[14]See Lichtenberg (1988b) for further discussion of the IR&D program.

[15]See Rogerson (1992a) for a more detailed discussion of unallowable costs.

5. ESTIMATION THEORY

This paper uses stock market data to estimate the size of the prize that firms were competing for. The basic idea is to calculate the stock market value of firms competing for a prime contract award a few days before it was announced which firm won and a few days after it was announced which firm won. Suppose there was no prize, i.e., the winner was expected to earn zero economic profit on the production contracts. Then the stock market value of the firms should not have been affected by the announcement. However, if the winning firm was expected to earn positive economic profit, its market value should have risen and the losers' should have fallen. Furthermore, the size of the changes in market value should be related to the size of the prize. The purpose of this section is to carefully outline a theory that describes the relationship between the observed changes in market value and the unobserved size of the prize. The theory will allow one to calculate the size of the prize associated with a given contest based on observation of the changes in market value.

A simple formal model of how the prize level and the contestants' probabilities of winning determine observed changes in market values will now be described. Since all the proofs are relatively straightforward they will not be included. Proofs are contained in Rogerson (1992b). The following notation will be used:

π	The dollar value of the prize that firms are competing for.[16]
n	The number of firms competing.
q	The probability as evaluated by the market that no firm will win and that the project will be cancelled.
p_i	The market's evaluation of the probability that firm i will win the contest conditional on the project not being cancelled. Thus, each p_i is between zero and one and the p_i's sum to one.
p	The vector of probabilities, $(p_1, \ldots, p_n)$.
p^*	The vector of probabilities, $(1/n, \ldots, 1/n)$.
V_i	The change in the market value of firm i on the day the winner is announced.
V_W	The change in the market value of the winner.
V_L	The sum of the change in the market value of the losers.
W	The index number of the winning firm.
$E(e/p,q)$	The expectation operator given p and q conditional on the project not being cancelled.

It will be assumed that investors are risk-neutral with respect to the risk of which (if any) firm will win the contract. This seems reasonable, since the contest risk is clearly idiosyncratic and thus diversifiable.

---

[16]The theory does not change if the possibility that different firms will receive different sized prizes is allowed for. It can be assumed that each firm i will receive a prize π_i if it wins, where the π_i are determined before the contest as independent draws from a distribution with mean π and the same formulas still apply.

An estimator of π is simply a real-valued function that maps the directly observable variables into an estimate of π. Formally, an estimator is therefore a function of $(V_1, \ldots, V_n)$ and W. Let $e(V_1, \ldots, V_n, W)$ denote an estimator. A good estimator is unbiased in the sense that it will on average estimate the size of the prize correctly. If an estimator is conservative in the sense that it tends to underestimate the size of π, this will also be useful, since a conservative estimator will establish a lower bound for π.

Definition: The estimator e is unbiased (conservative) given p and q if Eq. (5.3) [Eq. (5.2)] holds.

$$E(e(V_1, \ldots, V_n, W) \mid p,q) = \pi.\tag{5.1}$$

$$E(e(V_1, \ldots, V_n, W) \mid p,q) \leq \pi.\tag{5.2}$$

Finally recall that p and q are not directly observable. Therefore, it will be important to attempt to establish that an estimator is unbiased or conservative independent of p and/or q. Such estimators will be called uniformly unbiased or uniformly conservative.

Definition: An estimator e is

uniformly unbiased [conservative] over p given q	Eq. (5.1) [Eq. (5.2)] holds for every p
uniformly unbiased [conservative] over q given p if	Eq. (5.1) [Eq. (5.2)] holds for every q
uniformly unbiased [conservative] over p and q	Eq. (5.1) [Eq. (5.2)] holds for every p and q.

Consider the situation on the day before the prize is awarded. Firm i has a $(1 - q)p_i$ probability of winning the prize π. Thus, its market value is $(1 - q)p_i \pi$ higher than it otherwise would be.[17] All the losing firms' values will drop back down to zero above what they otherwise would be. This proves the following proposition.

Proposition 1: Suppose firm j wins. Then

$$V_i \begin{cases} (1 - p_j(1 - q)\pi & i = j \\ -p_i(1 - q)\pi & i \neq j. \end{cases}\tag{5.3}$$

An immediate corollary of this is the following.

Corollary 1: No matter which firms wins,

$$V_W + V_L = q\pi.\tag{5.4}$$

---

[17]By "otherwise would be" is meant the market value the firm would exhibit if it was not involved in the contest.

Therefore,

(i) $V_W > -V_L$ if $q > 0$.

(ii) $V_W = - V_L$ if $q = 0$.

There is a very natural intuitive explanation for Corollary 1. When $q = 0$, an announcement that a particular firm has won produces no news from the standpoint of the value of all n firms. It was already known that one of the n firms would win and this was incorporated into the market's estimate of the aggregate value of the firms. However, when $q > 0$, there is a chance that no firm will win and the project will be cancelled. Thus, an announcement that a particular firm has won produces new information from the standpoint of all n firms. The project will not be cancelled and one of the firms will receive π. Thus the aggregate value of all firms should increase from $(1 - q)\pi$ to π above what it otherwise would be, or equivalently, $V_W + V_L$ should equal $q\pi$.

The next step in determining a useful estimator of π is to calculate the expected value of V_W and $-V_L$. This is done in Proposition 2.

Proposition 2:

$$E(V_W/\, p,q) = \left[(1-q)\sum_{i=1}^{n} p_i (1- p_i) \right]\pi + q\pi. \qquad (5.5)$$

$$E(-V_L/\, p,q) = \left[(1- q)\sum_{i=1}^{n} p_i (1- p_i) \right]\pi. \qquad (5.6)$$

A useful estimator of π can now be constructed. Let $k(p)$ denote the function

$$k(p) = \frac{1}{\displaystyle\sum_{i=1}^{n} p_i (1- p_i)}. \qquad (5.7)$$

Then, define a class of estimators e_p as follows:

$$e_p = V_W + V_L - k(p)\, V_L . \qquad (5.8)$$

Proposition 3 describes its desirable property.

Proposition 3: e_p is uniformly unbiased over q given p.

Note that the only difference between the estimators in the e_p class is the size of $k(p)$. If $k(p)$ is larger, the estimator produces larger estimates of π. From Eq. (5.7), it is easy to see that $k(p)$ can be interpreted as a measure of the asymmetry of p. It is minimized if p is perfectly symmetric, i.e., $p = p^*$. (Recall that p^* is defined as $(1/n, \ldots, 1/n)$.) It equals infinity if p_i equals 1 for some i. This suggests two possible

approaches for constructing plausible estimates of π given that p cannot be directly measured. The first would be to use $e_p{}^*$ which always produces the smallest estimates within the e_p class. Proposition 4 shows that $e_p{}^*$ is in fact uniformly conservative over p and q.

Proposition 4:

(i) $k(p^*) = \dfrac{n}{n-1}$.

(ii) $e_p{}^*$ is uniformly conservative over p and q.

One possible approach for constructing a less-conservative, but still plausible estimator would be to assume that p was somewhat less symmetric. A natural assumption to make is that p is drawn from a uniform distribution, i.e., every p is equally likely. Let $E(e \mid u,q)$ denote the expectation operator conditional on the project not being cancelled given that p is distributed uniformly and for a given q.

Definition: An estimator e is said to be unbiased uniformly over q given u if Eq. (5.9) holds for every q.

$$E(e \mid u,q) = \pi. \tag{5.9}$$

Now define the estimator e_u by

$$e_u = V_W + V_L - k(u)V_L \tag{5.10}$$

where

$$k(u) = \frac{n+1}{n-1}.$$

Proposition 5 shows that e_u is unbiased in the above sense.

Proposition 5:[18] The estimator e_u is unbiased uniformly over q given u.

The theory developed in this section has three testable predictions. These are that $V_W \geq 0$, $V_L \leq 0$, and $V_W \geq -V_L$. It also provides a method for testing if $q = 0$. Finally, and most importantly, it also provides two rules for calculating π based on observed values of V_W and V_L. Sections 6 and 7 will now consider data from a number of major aerospace contests to test the predictions, test if $q = 0$, and calculate the size of prizes that were offered in these contests.

---

[18]The key step in the proof is showing the expected value of $\sum\limits_{i=p}^{n} p_i(1 - p_i)$ given that p is uniformly distributed and is equal to $1/k(u)$.

6. ESTIMATION OF CHANGES IN MARKET VALUE

Data were gathered for twelve major aerospace systems. The name of the programs, the date on which a winner was announced, and the identity of the winners and losers are presented in Table 3.1.

From the CRSP[19] tapes, daily percentage returns were created for three firms or artificial firms for each event. First, daily percentage returns for the winning firm were obtained. For the case of the F-18, a team of two firms won. Thus, the artificial firm consisting of both winners was created. For each event, this first firm is called the WIN firm. Second, daily percentage returns for the losing firm were obtained. When there were two losers, the percentage returns for the artificial firm consisting of both losers was created. For each event this second firm is called the LOS firm. Finally, daily percentage returns for the artificial firm consisting of all contestants were calculated. This third firm is called the TOT firm.

The standard event-study methodology was applied to estimate daily percentage returns corrected for marketwide movements.[20] The daily returns for the WIN, LOS, and TOT firms are graphed in Figures 3.1 to 3.3. Daily returns, which are significant at the 99 percent level, are circled.

Table 3.1

Description of Events

Program	Date	Winner	Loser
F-14	69/01/15	Grumman	McDonnell Douglas
S-3	69/08/01	Lockheed	General Dynamics
E-3A	70/07/08	Boeing	McDonnell Douglas
A-10	73/01/18	Fairchild	Northrop
F-16	75/01/13	General Dynamics	Northrop
UH-60	76/12/23	United Technologies (Sikorsky)	Boeing (Vertol)
KC-10	77/12/19	McDonnell Douglas	Boeing
A-7	64/02/11	LTV	North American Douglas
C-5A	65/09/30	Lockheed	Boeing, Douglas
F-15	69/12/23	McDonnell Douglas	Fairchild, Rockwell
B-1	70/06/05	Rockwell	Boeing, General Dynamics
F-18[a]	75/05/02	McDonnell Douglas Northrop	General Dynamics LTV

[a]For the F-18, McDonnell Douglas and Northrop competed as a team against the team of General Dynamics and LTV.

---

[19]See Center for Research in Security Prices (1986).

[20]See Rogerson (1992b) for details.
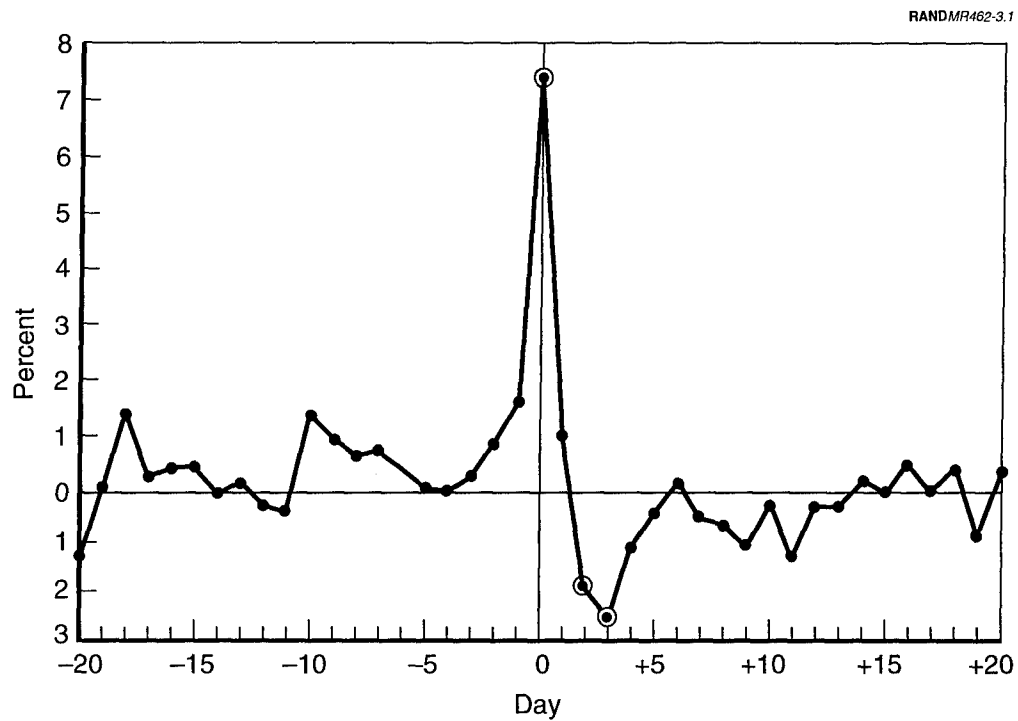
Figure 3.1—Excess Daily Returns for Winners
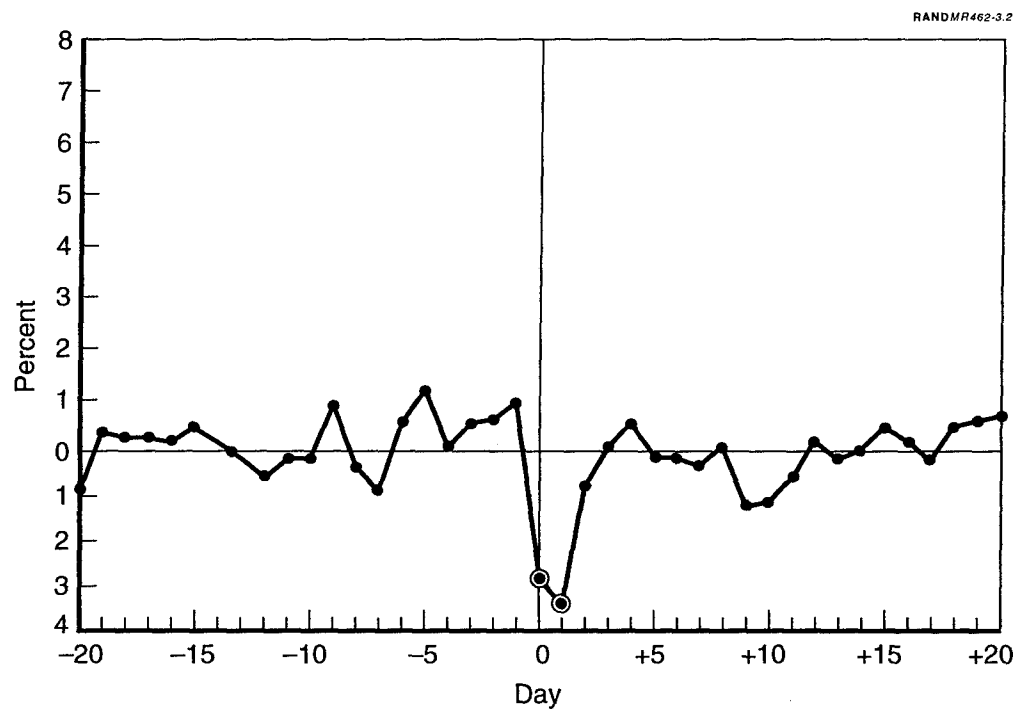
Figure 3.2—Excess Daily Returns for Losers
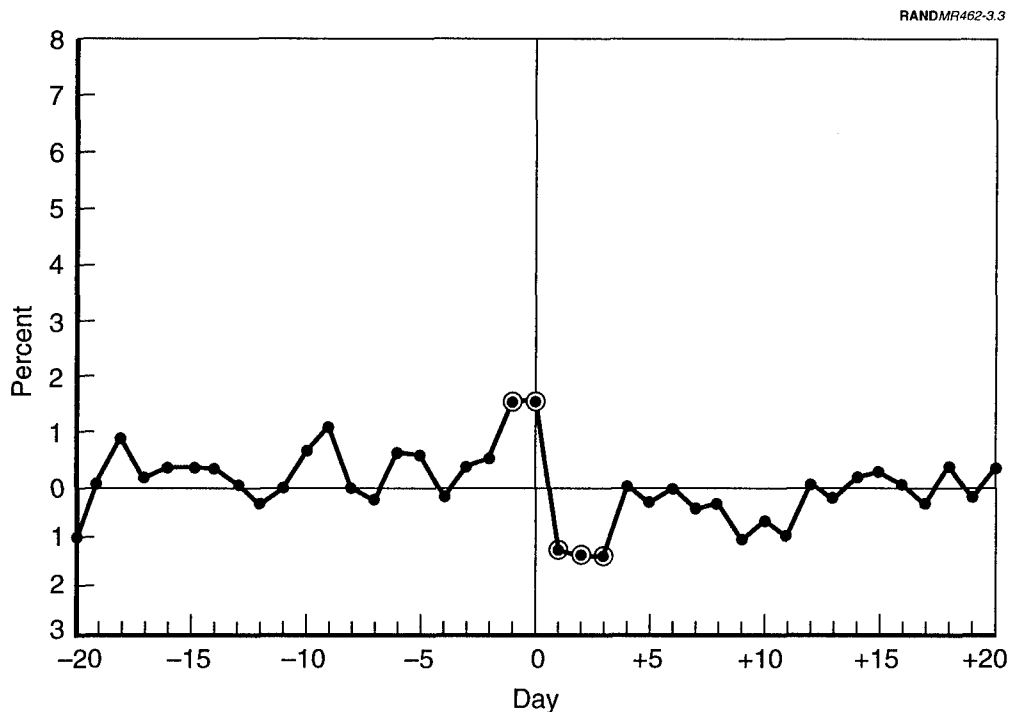
Figure 3.3—Excess Daily Returns for Total Firms

The above data can now be used to investigate the last major question that must be answered to calculate V_W and V_L. This concerns which day or days the announcement information was incorporated into the price of the stock. Define the "event window" to be the set of days on which the announcement information was incorporated into the stock price. Let $(\mu \ \upsilon)$ denote the event window beginning on day μ and running until day υ.

In theory, DoD's formal announcement procedure is to hold a press conference to announce its decision on the formal announcement day (day 0) *after* the market has closed for the day. Thus, news from this press conference should not affect the market value of firms until day 1. In practice, information often leaks out a day or two early. Furthermore, some details of the contractual arrangements the winner has agreed to are revealed a day or two later. Thus, some general "turbulence" in the returns might be expected for a few days before and after the announcement and the correct procedure would be to expand the event window to include these days.

Figures 3.1 to 3.3 suggest that expanding the event window to include two days before and after the announcement day is sufficient. Reference to Table 3.2, which presents the compound excess returns and associated t-statistics for various event windows, supports this. Thus, in the next section it will be assumed that the event window is (–2,3).

The value of V_W $(V_L, V_W + V_L)$ is calculated by multiplying the market value of the WIN (LOS, TOT) firm on the day before the event by the compound percentage

Table 3.2

Excess Returns and t-Statistics for Various Event Windows

μ	υ	WIN	LOS	TOT
		Excess Returns		
0	1	0.0851	−0.0601	0.0057
−1	2	0.0825	−0.0558	0.0100
−2	3	0.0656	−0.0499	0.0033
−3	4	0.0580	−0.0390	0.0081
−4	5	0.0545	−0.0394	0.0058
−5	6	0.0582	−0.0298	0.0128
−6	7	0.0594	−0.0276	0.0165
−7	8	0.0604	−0.0342	0.0113
−8	9	0.0548	−0.0469	0.0017
−9	10	0.0656	−0.0452	0.0089
−10	11	0.0701	−0.0505	0.0084
−11	12	0.0647	−0.0487	0.0102
−12	13	0.0618	−0.0547	0.0064
−13	14	0.0685	−0.0544	0.0085
		t-Statistics[a]		
0	1	9.556	−8.215	1.020
−1	2	6.682	−5.440	1.328
−2	3	4.527	−3.868	0.444
−3	4	3.552	−2.590	0.787
−4	5	3.003	−2.348	0.490
−5	6	2.856	−1.505	0.944
−6	7	2.691	−1.229	1.126
−7	8	2.629	−1.510	0.799
−8	9	2.340	−2.027	0.234
−9	10	2.557	−1.893	0.590
−10	11	2.543	−2.031	0.532
−11	12	2.296	−1.882	0.608
−12	13	2.103	−2.027	0.380
−13	14	2.201	−1.961	0.463

[a]$t_{.005} = 2.6$.

return of the WIN (LOS, TOT) firm over the event. Thus, one can test whether V_W (V_L, $V_W + V_L$) is positive, negative, or equal to zero by testing whether the percentage return of the WIN (LOS, TOT) firm is positive, negative, or equal to zero.

With reference to Table 3.2, the percentage changes of the WIN, LOS, and TOT firms are all strongly in accord with the theoretical predictions of the previous section. First, the percentage change of the WIN firms is significantly positive at the 99 percent level in accord with the prediction that V_W is positive. Second, the percentage change of the LOS firms is significantly negative at the 99 percent level in accord with the prediction that V_L is negative. Third, the percentage change of the TOT firms is positive but not significantly so. This agrees with the prediction that $V_W + V_L$ is nonnegative. Finally, since the percentage return of the TOT firms is not significantly different from zero, one cannot reject the null hypothesis that q equals zero.[21]

---

[21]Although it is not formally an unbiased estimator, one can rewrite Eq. (5.4) to obtain an estimator for q of $q = (V_W + V_L)/\pi$. Tables 3.3 and 3.4 will present, respectively, estimates for V_W and V_L and two estimates for π. Substituting these into the above equation yields estimates for q of between 0.004 and 0.005.

Table 3.3 presents the changes in the market value of the winners and losers for the event window (–2,3).

7. ESTIMATION OF PRIZES

Table 3.4 uses the estimator derived in Section 5 together with the estimates of changes in market value derived in Section 6 to calculate the size of prizes that firms were competing for in each event. The values labelled π_{LOW} and π_{HIGH} are, respectively, the values of the prize calculated using the lower (more-conservative) and higher (less-conservative) value of k.[22] Thus, the estimated average prize over all twelve contests is between $47 million and $67 million.

The natural question to ask is whether this is "large" or not. Two methods to answer this question are available. The first method is to compare the prize to the average market value of the competing firm. This is also done in Table 3.4. The row labelled

Table 3.3

**Estimated Change in Market Value of Winners and Losers
(in thousands of dollars)**

Event	V_W	V_L
F-14	8574	27613
S-3	22471	–9146
E-3A	8964	25751
A-10	6403	–13914
F-16	13592	–13013
UH-60	44727	–59619
KC-10	23836	–71616
A-7	3484	9547
C-5A	8910	–56014
F-15	42842	–56589
B-1	33128	–45062
F-18	102278	–54133
AVG	26601	–26350

Table 3.4

**Average Value of Prizes Across All Twelve Contests
(all dollar figures are in thousands of dollars)**

π_{LOW}	46,779
π_{HIGH}	66,957
M_B	457,249
π_{LOW} as a percentage of M_B	10.2%
π_{HIGH} as a percentage of M_B	14.6%
R	1,434,245
π_{LOW} as a percentage of R	3.26%
π_{HIGH} as a percentage of R	4.68%

---

[22]The prize was calculated for each of the twelve contests separately. The average of these twelve numbers is reported in Table 3.4.

"M_B" gives the average market value of the contestants on the day before the event (day –3). Then, the next two rows give, respectively, π_{LOW} and π_{HIGH} as a percentage of the average firm value. Thus, on average, the estimated prize is between 10.2 percent and 14.6 percent of the average value of the firms competing for it.

The second method is to compare the estimated prize to the expected discounted revenue stream that the winning firm will receive. The rationale behind this method requires a bit more explanation. Suppose that the winner will be involved in producing the weapon for T years. Let R_t denote the expected revenue that the winner will receive in year t of the project. Also assume that the owners of the firm use a discount rate of r to value these expected revenues. Let R denote the present discounted value of revenues that the winning firm will receive. This is defined by

$$R = \sum_{t=1}^{T} \frac{R_t}{(1+r)^t}.$$

(7.1)

Now suppose that the firm expects to earn α dollars of economic profit on every dollar of revenue it receives (where α is between 0 and 1). Thus, α is the economic profit rate that DoD is in fact allowing. The value of the prize to the winner, denoted by π, is defined by[23]

$$\pi = \sum_{t=1}^{T} \frac{\alpha R_t}{(1+r)^t}.$$

(7.2)

From Eqs. (7.1) and (7.2), it is clear that

$$\alpha = \frac{\pi}{R}.$$

(7.3)

That is, calculating the prize as a percentage of expected discounted revenues yields the economic profit rate that firms expect to earn.[24]

The value of π has of course been estimated. It remains to estimate R to calculate α. This was done by obtaining the DoD's estimate of the discounted expected cost of each system at the time of award[25] and correcting for two factors. These factors are as follows. First, not all of the cost of a weapon system is paid to the prime contrac-

---

[23]In reality α probably varies over the lifetime of the program. The calculation that is presented below can be viewed as the average value of α over the life of the program. In particular, note that profit margins may well be zero or even negative in the early years because the firm purposefully bid below cost on initial contracts to win the program or because the firm still spends some of its own money in full-scale development to insure that the program enters production. This means that the value of α on contracts signed after the choice of the prime may be higher than the number calculated below.

[24]The prize may in fact be larger than the RHS of Eq. (7.3) because part of the value of winning the contest may be that it becomes more likely that future contests will be won as well. In this case, the estimator in the RHS of Eq. (7.4) may be an overestimate of α.

[25]These were obtained from DoD documents called Selected Acquisition Reports (SAR) and the *Wall Street Journal.* The SAR data were reported in Dews et al. (1979).

tor. The cost includes money that is directly paid to firms that manufacture some major subcomponents of the system. One major such item for aerospace projects is the engine. The cost also includes the value of government-supplied equipment and services. No system-by-system data on the fraction of government cost that was actually received by the prime contractor as revenue could be located. However, a RAND case study of the F-16 contained data for the cost of the F-16 prototype phase, which showed that 68 percent of government's cost was paid to the prime contractors.[26] The second factor is that the discount rate used by the DoD in its cost estimates varied between 6.58 percent and 6.86 percent. Discount rates used by shareholders are likely to be much higher. To correct for these two factors, it was assumed that R is equal to 34 percent of the DoD cost estimate for each system. This corresponds approximately to 68 percent of the program cost being received by the prime and to shareholders using a discount rate double that of the DoD, i.e., approximately 13 percent or 14 percent.

The row in Table 3.4 labelled R presents the average value of R over the twelve systems and was calculated as described above. The next two rows present the estimate of α using, respectively, π_{LOW} and π_{HIGH}. Thus, every dollar of revenue received by a prime contractor on production contracts generates somewhere between 3.26 and 4.68 cents of pure profit.

---

[26]See Smith et al. (1981), p. 114. Engine manufacturing accounted for 27 percent of the total cost and government supplied equipment and services accounted for the remaining 5 percent.

WHEN MORE IS LESS: DEFENSE PROFIT POLICY IN A COMPETITIVE ENVIRONMENT
by Anthony G. Bower and Kent Osband[1]

1. INTRODUCTION

The Department of Defense (DoD) takes two different approaches to the assignment and pricing of procurement contracts. Some contracts are let competitively, with prices determined by sealed bids. Others are negotiated with sole-source suppliers, using so-called "profit policy" as a basis for price determination. Under profit policy, negotiators rely on past cost experience, as verified by audits, to estimate future expected costs.[2] Target price is intended to cover expected costs plus a target fee or "profit" equal to a percentage markup over expected cost. The markup varies by composition of outlays and assessment of performance and contract-type risk, according to the rules set forth in the "weighted guidelines."

If actual costs deviate from expected costs, the supplier on a negotiated contract may bear additional rewards or penalties. A firm-fixed-price (FFP) contract is closest in structure to a contract awarded by sealed bids: The supplier bears the entire burden of cost overruns and pockets the entire savings from underruns. Under a cost-plus-fixed-fee (CPFF) contract, the supplier is reimbursed its actual costs plus the target fee. Under cost-plus-incentive-fee (CPIF) or fixed-price-incentive (FPI) contracts, the supplier bears a constant fraction of cost underruns or overruns (subject to possible cost floors or ceilings).

For all the legal distinctions between competitively let and negotiated contracts, in practice the two contractual approaches often overlap. A competitively let contract may lead into a negotiated follow-on contract, or vice versa. Even a single contract may change midcourse, as for example, when design or delivery changes on a competitively let contract trigger negotiations over fair payment.

---

[2]Contractors are also required to submit cost estimates and to certify that the estimates are not intentionally fraudulent. The prospect of future auditing helps keep contractors from misreporting costs, but obviously there remains scope for strategic behavior.

The interaction between competition and negotiation is extremely complex. This paper focuses on one particular model within the broad theme of "competition to be a regulated monopolist." Our model assumes that an initial contract is let competitively to the low bidder, with common knowledge that the winner will later receive a single sole-source "follow-on" contract negotiated according to profit policy. Although stylized, the model is far from irrelevant to DoD practice. In 1986, competed contracts accounted for 57 percent of the $78 billion procurement budget, and follow-on contracts for 23 percent.[3] Price is rarely the sole consideration in contract award, but obviously it is a significant one. Moreover, its importance appears to be growing under political and budgetary pressures.

Our purpose is to study incentives within the context of current practice and determine the optimal contract within that framework.[4] The analysis is not particularly complex from a technical standpoint, although we do allow for both private bidding information (adverse selection) and discretionary learning (moral hazard). But the results run completely against the grain of popular thinking about cost-plus, negotiated contracts. To illustrate this claim, consider the following seemingly reasonable propositions about the model.

1. There is bound to be substantial inefficiency relative to the first-best full-information outcome.

2. Inefficiency is greater, the greater the procurement "weight" of negotiated contracts relative to competitively let contracts.

3. Total expenditures tend to rise with the size of the profit policy markup.

4. Higher markups are more lucrative for contractors.

5. Government negotiators can never benefit from using biased cost estimates.

Our analysis shows that each of these propositions is false. The factor that drives our results is that firms must bid for the right to have a profit policy contract later. The fundamental intuition is that positive profit policy markups serve to encourage more aggressive bidding. All firms try to "buy in" to the initial contracts, and higher-cost producers try to buy in more than lower-cost producers because they stand to receive larger absolute profit margins if they win. Hence, expected excess profit margins are shaved, and may drop to zero should the markup become sufficiently large.

The argument may be rephrased in terms of "handicapping." It is well known that when abilities can be observed, the more able competitors should be penalized relative to the less able. Under profit policy, proportional markups differentially subsidize the higher-cost producers. The handicapping tends to be self-correcting, except for the disincentive by-products of the auditing procedure used to identify cost. Also, higher markups discourage feasible cost reductions at the competitive stage of the project. To encourage cost reductions, the government should try to write the final

[3]Secretary of Defense (1988).

[4]For the unrestricted case and a comparison of the two models, see Bower (1993).

negotiated contract as soon as possible after the initial contract is written, or alternatively, to base negotiated price on a cost estimate that ignores first-period learning. In principle, either of these procedures can yield the first-best outcome.

Both in theoretical formulation and in the thrust of the analysis, this article is most closely related to work by McAfee and McMillan (1986). Like us, McAfee and McMillan model the competition for government contracts in the presence of moral hazard and adverse selection. Unlike us, they focus on the effect of cost-share parameters in a one-stage model. Auditing of costs occurs only at the very end of the procurement cycle. Nevertheless, they too identify a handicapping-type effect. Indeed, the present article was inspired by the realization that the reduced form of our two-stage model resembles the McAfee-McMillan structure.

Two articles by Laffont and Tirole (1986, 1987) address many of the same themes as McAfee and McMillan, in a slightly different context. McAfee and McMillan accept for modelling purposes the institutional restriction that the cost-share parameter is fixed before bidding. Laffont and Tirole derive optimal contracts when this restriction is dropped. They show that the optimal contract for risk-neutral suppliers is indeed linear in *ex post* cost but that lower-cost bidders should be induced to accept higher-cost shares.[5]

Our major purpose in studying this set of contracts is twofold: First, we bring together theoretical contracting literature with actual DoD practices. We demonstrate that at least one type of current DoD contract is closer to the theoretical optimal contract literature than previously recognized. This is not to claim that current DoD profit policy attains near-optimality. DoD policy appears to be directed as much toward regulatory notions of fairness as toward efficiency, without a clear focus on either. But, in this case at least, the DoD contract structure is capable of attaining nearly optimal results. Second, we draw several policy recommendations from our analysis. Thus, our paper focuses on bridging theory and practice rather than on extending the pure theory of contracts, although we do derive several new results.

The next section develops the basis model without moral hazard. Moral hazard is added in Section 3. Section 4 concludes.

2. THE MODEL WITHOUT MORAL HAZARD

Consider an industry composed of n risk-neutral, profit-maximizing firms. Each firm has a cost C, which is drawn independently from a distribution F with support $[C_{min}, C_{max}]$. F is common knowledge. The firm's cost is its private information (we avoid subscripting C by firm for later notational convenience). We assume that the firms have perfect information about their costs before bidding.[6] Each firm's reservation value, or opportunity cost, is set at zero. The government wishes to minimize pro-

---

[5]Rogerson (1987) provides some insights into why optimal contracts take this form.

[6]See Riordan and Sappington (1986) for a model in which firms gather information about their product costs over time.

curement expenditures for one unit of the procured good; quality is assumed to be set independently of contractors' bids and performance.

The game proceeds as follows. At time $t = 0$, firms bid via a second-price sealed-bid auction for an initial firm-fixed-price contract, which will last until time λ. Although a first-price auction is actually used in practice, we use a second-price auction because it yields simple, dominant strategies for the players. However, Myerson (1981) has shown that the expected revenues of a first- and second-price auction are identical if the bidders are risk neutral, their valuations are uncorrelated, and the seller (viz., the government) has a valuation sufficiently high. Our model satisfies these conditions, and hence we can use a second-price auction without loss of generality. After the bid at $t = 0$, the contract is awarded to the lowest bidder at the second-lowest bidder's bid. The winning firm produces the good continuously at a rate of 1 per unit time.

It is common knowledge that at time λ a sole-source follow-on contract lasting until time 1 will be issued to the winner of the first contract. There will probably be some design changes to the good; for convenience we will assume that the cost of the design changes is zero. (In general, assumptions of a determinate second contract can be relaxed without drastic effect on the model.) Price for the second contract will be calculated according to profit policy, as estimated cost plus a proportionate markup m. Estimated cost will be based on audited first-period cost, so payment will equal $(1 + m)C$ per unit time. We assume that the government can audit costs perfectly; we loosen this restriction at the end of Section 3. As shown in Laffont and Tirole (1986), the generalization to uncertain costs or a noisy monitor can be made easily. At time 1 the game ends. The game is played only once.

One may wonder why in practice the government does not simply write a single contingent contract covering all project work from time 0 to 1. There are several reasons why this does not happen. Full contingent contracts are difficult to write, especially for military procurements subject to rapid changes in both technological opportunities and assessments of need, which stochastically shift costs. As it is, major DoD contracts and supporting documents typically run to thousands or tens of thousands of pages. Even if contingencies could be spelled out, DoD officials might be unwilling or unable to commit to future actions because of lack of budgetary authority, fear of possible political repercussions, or uncertainty about future DoD policy.

An alternative view of the model is that the initial contract spans the entire project but that a design change triggering negotiation is anticipated at time λ. The old contract for the period λ to 1 will in effect be abandoned, replaced by a contract priced according to profit policy.

Denote C^i as the ith-lowest cost, and thus C^i represents the ith-order statistic from the cost distribution $F(C)$. Payments from the government are assumed to be made at a constant rate. This roughly corresponds to government practice for pure cost-plus-type contracts, but not for incentive contracts. On the latter, a portion of the costs, roughly 80 percent, is paid as the project proceeds (known as progress payments), with the remainder paid at project completion. This means that the present value of markups on incentive contracts tends to decline with project duration. In

particular, it is possible for a nominally positive markup to yield a negative present value (see Osband, 1988a). We will avoid this issue by assuming that all payments are converted into constant payment streams of equivalent net present value. Costs are assumed to be incurred at a constant rate.

Denote B as the firm's bid. Let B^i, $i = 1, \ldots, n$ be the ith-lowest bid. The firm's bid B will be interpreted as a flow of money at rate B per unit time. Both firms and the government discount at interest rate r, and the government commits to m before bidding. The winning firm's profit π is

$$\pi = \int_0^\lambda (B^2 - C)e^{-rt}dt + \int_\lambda^1 \left((1 + m)C - C \right)e^{-rt}dt.$$

To streamline the notation, we set

$$w_1 = \int_0^\lambda e^{-rt}dt = \frac{1}{r}(1 - e^{-\lambda r})$$

and

$$w_2 = \int_\lambda^1 e^{-rt}dt = \frac{1}{r}(e^{-r} - e^{-\lambda r}).$$

The variable w_1 is the present value of receiving \$1 per unit time from time 0 to time λ, and w_2 is the present value of receiving \$1 per unit time from time λ to time 1. If $r = 0$, w_1 and w_2 simply equal the lengths λ and $1 - \lambda$ of the respective two periods. Thus, the profit equation for the winning firm is $\pi = w_1 (B^2 - C) + w_2 mC$, which can be rewritten as

$$\pi = w_1 \left(B^2 - \left(1 - \frac{w_2}{w_1} m \right) C \right). \tag{1}$$

Total government expenditures x are profits plus costs to the firm, or $x = w_1 B^2 + w_2 (1 + m)C$.

Observe that x is a linear combination of the bid and cost C, and thus this two-period contract has been reduced to a one-period form, as in McAfee and McMillan (1986). McAfee and McMillan assume a contract form $P = (1 - \alpha)B + \alpha C$, with the cost share $\alpha < 1$ (i.e., the firm faces an incentive contract). In our model this corresponds to the condition $w_2(1 + m) < w_1 + w_2$, or $m < w_1 / w_2$. From inspection of Eq. (1), we can see that if $m < w_1 / w_2$, then the firm faces an incentive contract. Thus, at this juncture we could appeal directly to McAfee and McMillan, Theorem 2, for our results. However, to aid the intuition, it is worthwhile to proceed a bit further with the analysis of the bidding strategies of the firms.

Our use of a second-price auction facilitates the analysis, because a winning firm's profit on the contract is independent of its bid. It follows that each firm i has a dominant strategy of bidding its true cost, or

$$B^{i=}\left(1 - \frac{w_2}{w_1} m\right)C^i. \tag{2}$$

(Substitute Eq. (2) into Eq. (1) to verify that this is the equilibrium.) The lowest-cost producer receives the contract, and it receives the second-lowest virtual cost $[1 - (w_2 / w_1)m]C^2$ in payments.

From the government's perspective, the expected profit, *ex ante* (i.e., before any firm knows its cost), of the winning firm is $E[\pi] = (w_1 - w_2 m)(E(C^2) - E(C^1))$, where, again, C^1 and C^2 represent the first- and second-order statistics of F. Note that profits decrease as a function of m. This is the key insight of our model. In McAfee-McMillan, this is called the "bidding competition" effect. Evidently, the government can overcome the adverse selection of unknown costs by correctly designing the auction and contract. In Eq. (1) the term $w_2 mC$ represents second-contract profit and is increasing in C. Thus, higher-cost firms make more second-contract profit than low-cost firms, which offsets some of the cost advantage of the low-cost firm. Each firm takes this into account when bidding for the contract. Hence, high-cost firms bid more aggressively with higher markups, which shaves the expected profits of the low-cost firms.

Define $m^\lambda = w_1 / w_2$. Expected government expenditures are

$$E(x(m)) = w_1 E(B^2) + w_2 (1 + m)E(C^1)$$
$$= w_1 E(C^2) + w_2 E(C^1) - mw_2 (E)C^2) - E(C^1)), \quad \text{if} \quad m < m^\lambda.$$

If $m = m^\lambda$, firms bear no cost to produce the good, and so all firms bid zero. In that case, the government cannot identify the lowest-cost producer and pays in expectation $E(x) = (w_1 + w_2)E(C)$. Thus, it is optimal to impose some cost strictly greater than zero on the firm. The government minimizes total expenditures by setting m as high as it can (within, say, $\in$ of m^λ) while still selecting the low-cost firm.

The results are summarized in the following proposition. See Figure 4.1.

Proposition 1: Given private information about costs but no moral hazard, the government can approximate arbitrarily closely the first-best solution by setting $m = w_1 / w_2 - \in$, with $\in$ arbitrarily small. The government pays the lowest-cost firm (slightly over) its cost, and the firm makes (almost) zero profit.

Our result is closely related to a result on auctions by Riley (1988). Riley examines a pure "oil-lease" auction in which a signal that is correlated with the buyer's valuation of the object, such as the number of barrels of oil obtained from the leased land, is observable *ex post*. In our model, the signal corresponds to audited cost and the buyer's valuation corresponds to the firm's initial cost C. Riley shows that seller revenues increase in the royalty rate, provided that buyer profits are monotonic increasing in the signal. In our model, the equivalent statement is that government expenditures decrease in m, provided that profits decrease in C—or, in other words,

RAND*MR462-4.1*
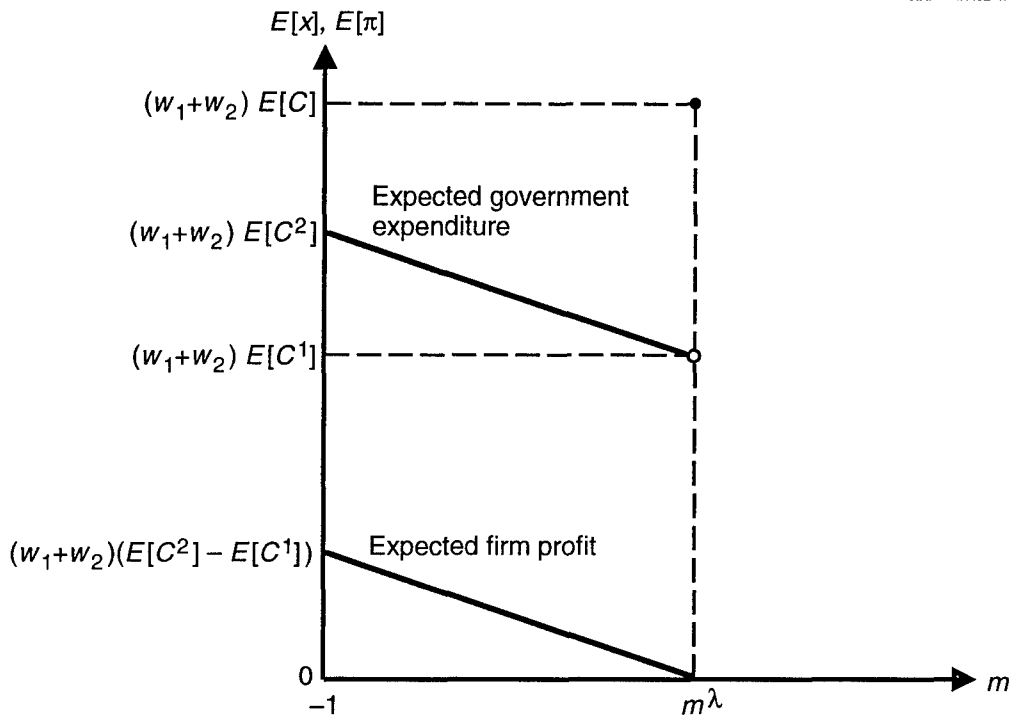
Figure 4.1—Government Expenditure and Firm Profit
(without moral hazard)

provided $m < m^\lambda$. The markup m plays a similar role as the royalty rate. Note that expected expenditures jump up at exactly the point that monotonicity of profit in audited cost is violated.

It is curious how a seemingly inefficient cost-based reimbursement scheme can provide a tool for cost reduction. The key is the interaction with competition. Cost-based reimbursement helps the weakest (highest-cost) firms the most, yet it need not help them so much that they actually win the contract. The auction then recaptures the excess profits from a high markup. Hence, pure auctions and pure cost-plus contracts are outperformed by a hybrid of the two. McAfee and McMillan and Laffont and Tirole both consider linear incentive schemes and find that incentive contracts, with some cost sharing by the government, are optimal. Our result is closely related to theirs, because our markup m plays the same role as their "cost-share parameter": It provides a way to reimburse firms for their costs. The result is perhaps more striking in this context because it seems at first glance to be counterintuitive. Of course, by now we hope to have conveyed the correct intuition.

Another striking feature of the model is the "knife-edge" character of the result. For the government, cost minimization lies within ϵ of cost maximization. Hence, if there is uncertainty about the relative weights w_1 and w_2, regulators should err on the low side rather than the high in the choice of m. As we shall see, knife-edge phe-

nomena recede in importance once cost-reducing effort is introduced into the model.

Observe that at the optimum, firms buy in to the first contract with a lowball bid and then recover their outlays on the second contract via a generous markup. If DoD cannot be trusted to follow through on the second contract, this is not a viable solution.

Moreover, even a DoD that intended to implement optimal contracts might be prevented from doing so. The extreme variability of payment flows from contract to contract is almost bound to invite political intervention, as interest groups on each side try to adjust contract provisions in their favor. As Baron (1988) has shown, *ex post* corrective action by lobbies is likely to reduce long-run expected welfare. In our model, as in Baron's, lobbying tends to be counterproductive for the lobby itself. Restrictions on maximum contract markups—typically favored by public "watchdog" groups—can serve to raise long-run expected profits, while floors on minimum markups—presumably favored by contractors—can serve to remove profit entirely. These results arise directly from our earlier counterintuitive result that profit and government expenditures are often decreasing in the markup.

3. THE MODEL WITH MORAL HAZARD

We now introduce the possibility of unobservable cost-reduction effort by the firm. Not unexpectedly, profit policy will affect the equilibrium level of effort. High markups will tend to retard cost reduction. Hence, the presence of moral hazard can reduce the optimal markup, as the government trades off decreased information rents to firms in the bidding phase with increased realized cost reductions during the production phase.

We will construe effort as an investment in permanent improvements in production technique ("learning"), so that costs fall with cumulative effort. Specifically, we will assume that a firm that exerts effort $\eta_i(\eta_i \geq 0)$ over a period dt will lower its costs permanently by $\eta_i \bullet dt$. Note that this cost reduction is independent of the firm's cost. Effort imposes a disutility on the firm of $\psi(\bullet)$, with the usual assumptions on $\psi(\bullet)$: $\psi(0) = 0, \psi' > 0$, and $\psi'' > 0$. In particular, the marginal disutility of effort increases with effort. To ensure that the second-order condition for a solution is satisfied, we will need the further assumption that $\psi''' \geq 0$.[7]

In this section we will assume that the firm has only two opportunities to set effort: It sets η_1 at the start of the first contract $(t = 0)$ and η_2 at the start of the second contract $(t = \lambda)$. The time path of costs is as follows. From $t = 0$ to $t = \lambda$, costs drop linearly from C^1 to $C^1 - \lambda\eta_1$. From $t = \lambda$ to $t = 1$, costs drop steadily from $C^1 - \lambda\eta_1$ to $C^1 - \lambda\eta_1 - (1 - \lambda)\eta_2$. Total first-period cost C_1 to the firm is

$$C_1 = \int_0^\lambda (C^1 - \eta_1 t)e^{-rt}dt = w_1 C^1 - \upsilon_1\eta_1, \tag{3}$$

---

[7]This assumption is also used in Laffont and Tirole (1986, 1987).

where $\upsilon_1 \equiv (1/r)(w_1 - \lambda e^{-\lambda r})$. The coefficient υ_1 represents the present value of direct first-contract cost reductions on first-period cost. The government observes average first-period cost $C^1 - (\lambda/2)\eta_1$ and bases second-period revenues on it. Total second-period cost C_2 is

$$C_2 = \int_\lambda^1 \left(C^1 - \eta_1 \lambda - \eta_2 (t - \lambda) \right) e^{-rt} dt = w_2 (C^1 - \eta_1 \lambda) - \upsilon_2 \eta_2, \tag{4}$$

where $\upsilon_2 \equiv (1/r)(w_2 - \lambda(1 - e^{-\lambda e}) - e^{-r})$. The coefficient υ_2 represents the present value of direct second-period cost reductions on second-period cost. Profit is

$$\pi = w_1 B^2 - C_1 + w_2 (1 + m)(C^1 - \lambda \eta_1 / 2) - C_2 - w_1 \psi(\eta_1) - w_2 \psi(\eta_2). \tag{5}$$

Substituting Eqs. (3) and (4) into Eq. (5) yields

$$\pi = w_1 \left(B^2 - \left(1 - \frac{w_2}{w_1} m \right) C^1 \right) - w_2 (1 + m)\eta_1 \lambda / 2$$
$$+ \upsilon_1 \eta_1 + w_2 \eta_1 \lambda + \upsilon_2 \eta_2 - w_1 \psi(\eta_1) - w_2 \psi(\eta_2). \tag{6}$$

All terms of the profit equation are measured in present value. The first term represents the bid minus the cost incurred in the first period, plus the second-period profits from the profit policy. The second term is the second-period revenue loss from first-period cost reduction. The third and fourth terms are the first-period and second-period gains from first-period cost reduction, respectively. The fifth term is the second-period gain from second-period cost reduction, and the sixth and seventh terms are the direct costs of first- and second-period effort, respectively. The firm controls η_i and η_2 and maximizes profit. Note that the moral hazard terms are linearly separable from the cost terms—there is no interaction between initial cost and effort.

Differentiation of Eq. (6) with respect to η_i and η_2 establishes that the profit-maximizing level of effort η_1^* and η_2^* satisfy $\psi'(\eta_2^*) = \upsilon_2 / w_2$ and

$$\psi'(\eta_1^*) = \frac{\upsilon_1}{w_1} + \lambda \frac{w_2}{2w_1}(1 - m). \tag{7}$$

Note that second-period effort is set at the first-best level υ_2 / w_2 because the firm gets to keep all of its cost reductions in the second period. In the public utility literature, this is known as a "regulatory lag" effect. First-period effort is distorted downward because of moral hazard. The firm faces the constraint that $\eta_i \geq 0$, thus $\psi'(\eta_i) \geq 0$. This implies that for all $m > m^0 \equiv 2\upsilon_1 / (\lambda w_2) + 1$, $\eta_i = 0$. This must be taken into account in the government's cost-minimization problem.

Profits in Eq. (6) can be written in a simpler form as

$$\pi = w_1 B^2 - (w_1 - w_2 m)C^1 + V^*(m),$$

where $V^*(m)$ is the net gain from effort and represents everything after the first term in Eq. (6), with effort in periods 1 and 2 set optimally, or

$$V^*(m) = -w_2(1+m)\frac{\lambda}{2}\eta_1^*(m) + \upsilon_1\eta_1^*(m) + \lambda w_2\eta_1^*(m)$$

$$+\upsilon_2\eta_2^* - w_1\psi(\eta_1^*(m)) - w_2\psi(\eta_2^*).$$

Since $V^*(m)$ does not vary from firm to firm (here is where the assumption of cost reductions being independent of previous cost becomes key), these net gains get bid away by the firms and entirely captured by the government.

Lemma: Given adverse selection with linearly separable moral hazard, a firm with cost C will bid $[1 - (w_2/w_1)m]C - V^*(m)/w_1$. Profits are the same as in a pure adverse-selection model; all cost-reduction gains are captured by the government.

However, the size of those gains depends inversely on m, so the government faces a tradeoff between increased bidding competition versus increased first-period cost reduction. To solve for the optimal m, we start by writing expected government expenditures as

$$E(x(m)) = w_1 E(B^2) + w_2(1+m)\left(E(C^1) - \frac{\lambda}{2}\eta_1^* m\right)$$

$$= (w_1 - w_2 m)E(C^2) - V^*(m) + w_2(1+m)\left(E(C^1) - \frac{\lambda}{2}\eta_1^*(m)\right). \qquad (8)$$

For now, assume that the constraints $m \le m^\lambda$ and $\eta_1 \ge 0$ do not bind. The derivation of the first-order condition for m is simplified by using the envelope theorem to derive

$$\frac{\partial V^*(m)}{\partial m} = -\frac{\lambda}{2} w_2\eta_1^*(m).$$

Taking the derivative of Eq. (8) with respect to m establishes that

$$E(C^2) - E(C^1) = -\frac{1}{2}(1 + \underline{m})\lambda\frac{d\eta_1^*(m)}{dm}, \qquad (9)$$

where $\underline{m}$ denotes the optimal markup, which is assumed to be an interior stationary point. To determine $(d\eta_1^*)/(dm)$, differentiate the equilibrium condition (7) for η_1^* to obtain

$$\psi''(\eta_1^*)\frac{d\eta_1^*}{dm} = -\frac{\lambda w_2}{2w_1}. \qquad (10)$$

Combining Eqs. (9) and (10) yields an implicit expression for $\underline{m}$:

$$\underline{m} = \psi''(\eta_1^*)\frac{4w_1}{\lambda^2 w_2}(E(C^2) - E(C^1)) - 1. \tag{11}$$

It is easily checked that a sufficient condition for the second-order condition to be satisfied is $\psi''' \geq 0$. Equation (11) identifies the optimal m given that the constraints do not bind. However, if $m > m^\lambda \equiv w_1/w_2$, then firm profits will increase in C, as can be seen from the inspection of the first term of Eq. (6). Thus, it is still the case that m is constrained to be less than m^λ. Also, if $m \geq m^0 \equiv (2u_1)/(\lambda w_1) + 1$ (i.e., the first-order condition on first-contract effort requires $\eta_1 < 0$ to be satisfied), the firm applies zero effort in the first period. The effects of the constraints are taken into account in the optimal profit policy, which is outlined in Proposition 2. The proof involves straightforward checking of the constraints and is omitted. The three key properties are first, if $m^\lambda \geq m \geq m^0$, then government expenditures are decreasing in m; second, if $m^0 \geq m \geq m^\lambda$, then government expenditures are increasing in m; and third, if $m \geq m^0$ and $m \geq \max\{m^0, m^\lambda\}$, then government expenditures are constant in m.

Proposition 2: Let $m_.^\lambda$ denote the closest feasible markup to m^λ that is less than m^λ. Under moral hazard and adverse selection the government minimizes expenditures at

$$m^* = \begin{cases} m_.^\lambda & \text{if} \quad \underline{m} \geq \min\{m^0, m^\lambda\} \\ \underline{m} & \text{if} \quad \underline{m} \leq m^\lambda \leq m^0 \\ \underline{m} & \text{if} \quad \underline{m} \leq m^0 \leq m^\lambda \quad \text{and} \quad x(\underline{m}) \leq x(m^\lambda) \\ m_.^\lambda & \text{if} \quad \underline{m} \leq m^0 \leq m^\lambda \quad \text{and} \quad x(m^\lambda) < x(\underline{m}). \end{cases}$$

If the interest rate is zero, then $m^0 = (2 \bullet \lambda^2/2)/[\lambda(1-\lambda)] + 1 = \lambda/(1-\lambda) + 1 > m^\lambda$ and Proposition 2 simplifies to a corollary given below.

Corollary: If $r = 0$, then the government minimizes expected expenditures at

$$m^* = \min\{\underline{m}, m_.^\lambda\}. \tag{12}$$

Figures 4.2 and 4.3 illustrate the corollary, which recalls similar results by McAfee and McMillan (1986). Figure 4.2 shows an example where $\underline{m} < m^\lambda$ and Figure 4.3 shows the opposite. In Figure 4.2, government expenditures fall with m at a decreasing rate down to $\underline{m}$ and then increase, whereas profits drop linearly to zero at m^λ. Again, the intuition is that the firm earns profits only from its information rents, passing on all cost-reduction gains (or lack of gains) to the government. In the region $[\underline{m}, m^\lambda]$, government expenditures rise even as profits fall, because of the severe moral hazard effects. The marginal increase in government outlays for firm profits from lowering m is more than offset by the marginal increase in cost reduction gains. Thus, in this region the preferences of the two parties regarding m coincide, and the government decreases net outlays by stimulating cost reduction and allowing some profits. The government allows the winning firm positive profits in equilibrium to

$E[x], E[\pi]$
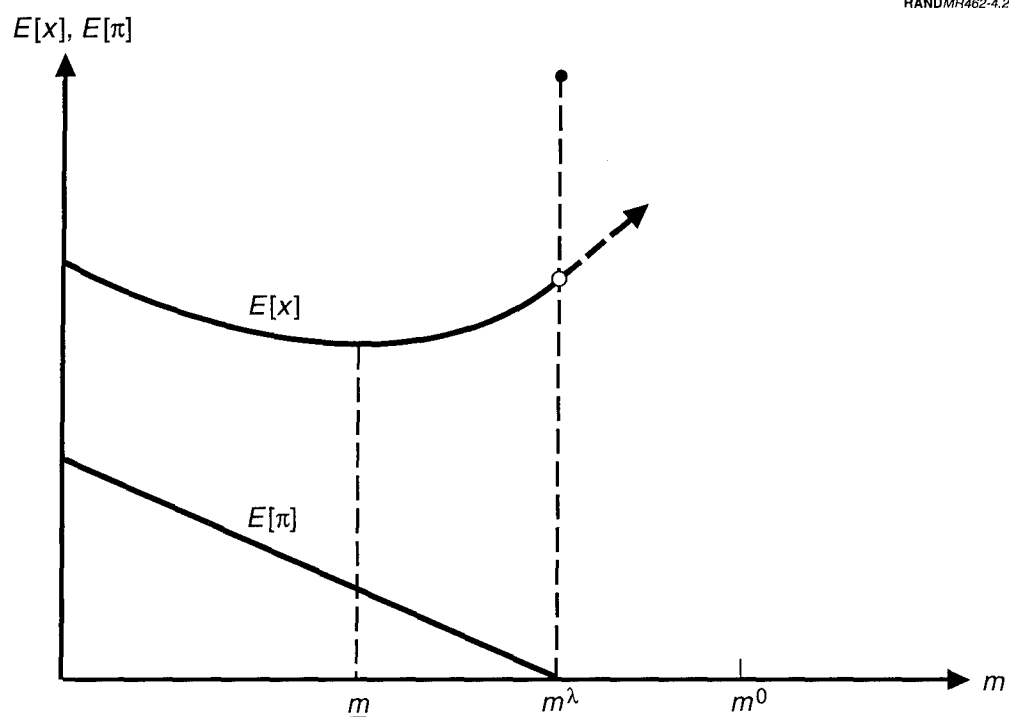
Figure 4.2—Government Expenditure and Firm Profit
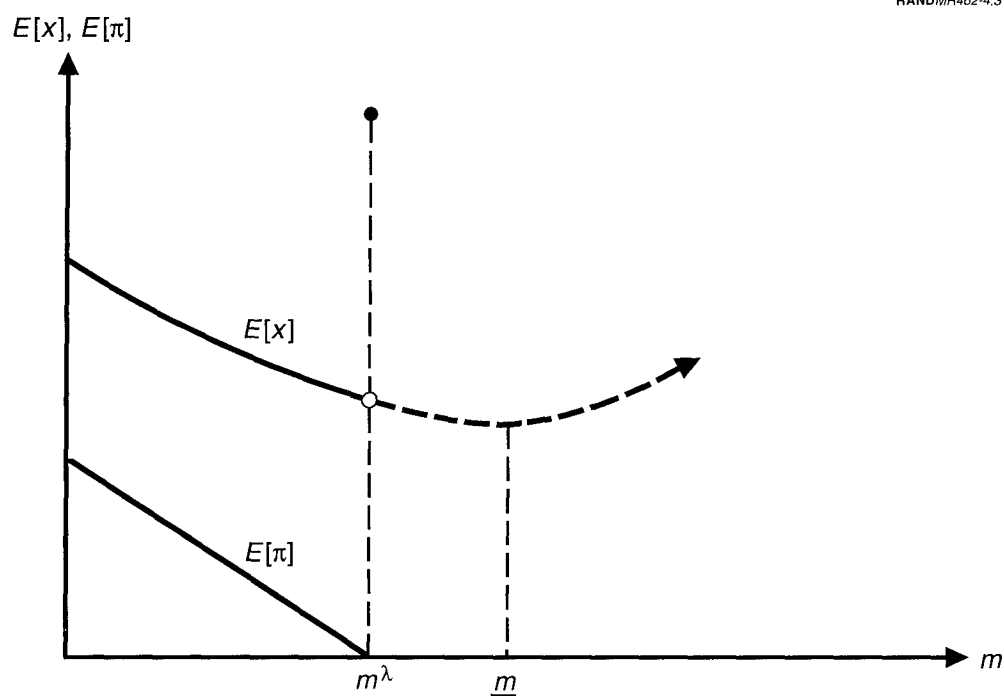(with moral hazard and $\underline{m} < m^{\lambda}$)

$E[x], E[\pi]$

Figure 4.3—Government Expenditure and Firm Profit
(with moral hazard and $\underline{m} > m^{\lambda}$)

avoid strong moral hazard effects. In Figure 4.3, the government always gains more from reducing profits than it loses from withheld effort, and a corner solution is obtained.

To gain more intuition as to when a corner solution is obtained, consider the case of $r = 0$ and quadratic disutility $\psi(\eta_i) = (K/2)\,\eta_i^2$. Then, from Eq. (11),

$$\underline{m} > m^\lambda \iff E(C^2) - E(C^1) > \frac{\lambda}{4K}.$$

Thus, if marginal losses to adverse selection, proportional to $E(C^2) - E(C^1)$ and constant in m, are larger than marginal losses to moral hazard at $m = m^\lambda$, which are increasing in m and equal to $\lambda/4K$ at $m = m^\lambda$, then set $m = m^\lambda$ and drive firm profits to (almost) zero.

Still assuming quadratic disutility ($\psi'' = K$), a number of comparative-statics results can be gleaned from inspection of Eq. (11). The markup $\underline{m}$ is increasing in the expected size of the pure information rents $E(C^2) - E(C^1)$. One possible way to shrink $E(C^2) - E(C^1)$ is to increase the number of capable firms bidding. The more firms bidding, the smaller the markup can be, because competition is mitigating the information rents. In principle, markups should be set on a sector-by-sector or even a contract-by-contract basis, to adjust to varying technological uncertainties, cost-reducing prospects, and industry competitiveness. The trend of the last quarter century has been exactly the opposite: to standardize markup policy. This may have advantages from the perspective of internal government monitoring and avoidance of congressional oversight, considerations that our model ignores. But direct contracting efficiency appears to be sacrificed. Finally, the markup is directly proportional to K; a high K implies that changes in markups will have smaller marginal influence on first-period effort η_1^*.

Biased and imperfect auditing. The model can be generalized to allow for different auditing procedures and for cost-sharing on the second contract. Our auditing baseline assumed that the government observes total cost on the first contract and uses that as an estimate of second-contract cost. Now allow auditors to observe costs at different times within the first contract and to use that information to form an estimate $C^1 - s\lambda\eta_1 - S$ of average second-contract cost. The parameter s represents the proportion of first-period cost reductions included in the estimate, and S represents a lump sum, possibly an *a priori* estimate of second-contract cost. For example, the baseline can be expressed as $s = 1/2$, $S = 0$. If $s = 1$ and $S = ((1 - \lambda)/2)\,\eta_2^*$, the auditor will have an unbiased estimate of costs for the second contract. Estimates for which s is less than 1 will be called "lagged" because they are historical measures and overestimate future cost, whereas estimates for which s exceeds 1 will be called "leading." We also allow for the government to bear a share $1 - \theta$ of second-contract cost underruns or overruns. The parameter θ is called the (contractor's) cost share, and it will be allowed to vary between 0 and 1. Our baseline case sets θ equal to 1.

Observe that S has no effect on choice of effort, firm's profit, or total government expenditures because it is simply a lump-sum payment that does not vary from firm to

firm. S is completely bid away in the auction in a fashion similar to the efficiency gains $V^*(m)$. The remaining parameters s and θ change the first-order conditions for effort to

$$\psi'(\eta_1^*) = \frac{\upsilon_1}{w_1} + \lambda \frac{w_2}{w_1} (\theta - \theta s - ms) \tag{13a}$$

and

$$\psi'(\eta_2^*) = \theta \frac{\upsilon_2}{w_2}. \tag{13b}$$

Equation (11) for $\underline{m}$ is replaced by

$$\underline{m} = \frac{\psi''(\eta_1^*)w_1}{s^2\lambda^2 w_2} \left(E(C^2) - E(C^1) \right) - 1. \tag{14}$$

From inspection of Eq. (13b) we see that second-contract effort increases with θ and that second-contract effort is independent of s. Also, from Eq. (13a), first-contract effort is decreasing in s. It follows that a lagged estimate $(s < 1)$ is superior to an unbiased or leading estimate, since the government reduces total outlays through stimulated cost-reduction effort. The optimal $\underline{m}$ increases as s falls, until eventually the constraint $m \leq m^\lambda$ binds. The efficiency loss from the constraint diminishes as s approaches zero.

In the limit, with $s = 0$ and $\theta = 1$, incentives for cost-reducing effort are first-best in each contract, as can be seen from Eqs. (13a) and (13b). Optimal effort is higher in the first contract than the second, since all first-contract learning carries over to the second contract but not vice versa. When $s = 0$ in the limit, cost estimates no longer depend upon variables subject to moral hazard. Manipulation of the markup squeezes the information rents out of the winner of the auction, while the two firm-fixed-price contracts are independent and impose no effort penalty. We do not claim that such careful measurement is feasible—in particular, for the government to use a particular s in its cost estimation, it is necessary that costs be observable at time λs. For small s this will not be feasible. But the result clearly illuminates the underlying logic of the model and the pivotal role of audit timing. Results are summarized in the following proposition.

Proposition 3: Efficiency is improved by using lagged estimates of second-period costs and higher cost shares. As lag and cost share move to $s = 0$, $\theta = 1$, procurement cost in the limit approaches the first-best.

One more generalization concerns the possibility that monitoring is imperfect. Suppose that audits measure first-contract costs (at whatever time) plus additive random disturbances. The disturbance terms will enter as separate terms in the firm's profit

equation (Eq. (6)). Since the firm is interested in maximizing expected profit, the random disturbance terms will drop out. Provided those disturbances are independent of true costs, neither the noisiness of the measures nor possible biases will affect effort choice or expected overall payments. Thus, if firms are risk neutral, then a linear contract is optimal and the Laffont and Tirole (1986) result holds here as well. Of course, for large disturbances, risk neutrality may not be a good assumption.

4. CONCLUSIONS

Government expenditures generally decrease in m for small m, because of bidding competition effects. The optimal markup m^* is low only when the distribution of costs is tight (i.e., the information on how much it costs to make the product is very good); there is a low disutility of effort; and the first contract is short.

Firms will buy in to a contract with a low bid, expecting to make up their loss during the cost-plus stage. The higher the m, the lower the bid. Bids are very sensitive to m, but government expenditures are less sensitive. Bids represent first-period wealth, and the size of m has a first-order effect on first-period wealth as higher m moves wealth from the first period to the second period; m has only a second-order effect on the overall amount of wealth as it retards first-period effort. The effect m has on revenue distribution over time is particularly large on long-term contracts, since the component of the bid V^*/w_1 gets large as w_1 shrinks. Once the correct economic institutional features are in place, namely, the auction and some sort of reimbursement for costs, government expenditures appear to be rather insensitive to m over fairly broad ranges. Firms will bid away excess second-period profits during the auction. Thus, in a loose sense, having the feature of a competitive auction is more important than the exact level of profit allowed in the second period, because the information-eliciting auction provides the safety net of competition to the government.

This model can be extended in many ways. An extension examined by Bower (1993) is to place the model in the general, optimal-contracting framework of Laffont and Tirole. In their model, the cost-share parameter is allowed to vary with the bids; in our model, m does not depend upon the bid. Bower shows that the cost of this restriction to DoD is usually quite small. He also examines the case of renegotiation at the end of the first period and an extension to two or more auditing periods.

BUY-IN AND GOLD PLATING UNDER DEFENSE PROFIT POLICY
by Yeon-Koo Che[1]

1. INTRODUCTION

Two institutions distinctively characterize the Department of Defense's (DoD)'s weapon systems procurement. One is competitive source selection; the other is a set of regulations called "profit policy," which covers noncompetitive, negotiated contracts. Despite the fact that many contracts are initially awarded through competitive source selection,[2] eventually most become subject to profit regulation at the recontracting stage, where the DoD frequently finds itself compelled to negotiate with an incumbent contractor on a sole-source basis because of the latter's project-specific skills collected during the initial contracting period.

The essence of profit policy is cost-based pricing.[3] A rather complicated rule, called "weighted guidelines," determines the target profit as a percentage markup over the expected costs. The cost estimates are in turn calculated from the historical costs.

In this paper, I study how this profit policy regulation affects the performance of defense contractors, especially in determination of quality variables. To this end, I adopt a stylized two-period scenario, similar to that considered by Bower and Osband (Chapter Four), where a competitively let contract leads to a sole-source follow-on contract that is regulated by profit policy. My model is distinct from theirs in that the first-period competition involves a nonprice dimension. Specifically, I model it as a two-dimensional auction, where each firm bids on both price and design specification (hereafter called "quality"). Unlike the traditional approach, which

---

[2]In 1986, the competitively let contracts accounted for 57 percent of total procurement (Secretary of Defense, 1988).

[3]This is in contrast to the regulation of public utilities where profitability of a utility is linked to capital assets employed. However, a recently revived markup consideration over the facility capital in defense procurement signals that profit policy is moving closer to the rate of return regulation. For a detailed description of profit policy, see Osband (1989a) and Rogerson (1992a).

assumes that competition is confined to the price dimension, this approach reflects the current practice adopted in many government competitive procurements, where heterogeneous designs compete actively.[4] Che (1993) establishes general theoretical results about two-dimensional auctions, which this paper will refer to.[5]

Absent the quality consideration, this model closely resembles those of Bower and Osband (Chapter Four) and McAfee and McMillan (1986). In the similar setting, Bower and Osband (Chapter Four) found that a positive markup in the second-period profit policy regulation can serve as a discriminatory bidding subsidy in the first-period auctions that favors a high-cost more than a low-cost firm. This bidding subsidy is desirable for the buyer, since it makes a low-cost-type firm bid more aggressively in the first-period competition. Such a "bidding competition" effect can be fully utilized by raising the markup to the point where effective cost differentials across types approach zero.

Adding a quality dimension in this framework makes the effect of markup more complex. When quality is endogenously determined as part of the competitive process, firms enjoy an additional strategic leverage. Now, firms can dilute the bidding competition effect by increasing quality. Increasing quality in the bid not only improves a firm's chance of winning but also increases the regulated profits that the firm would earn in the second period should it win the competition. Therefore, raising a markup has an undesirable side effect of creating excessive design competition, which results in a quality distortion. This result is reminiscent of the "Averch-Johnson effect" (Averch and Johnson, 1962), which first showed that rate of return regulation leads to excessive capital for a reason similar to one developed here. In our context, the Averch-Johnson distortion can arise as a result of the buyer's effort to elicit the bidding competition effect.

A crucial assumption of the model is that the buyer is myopic and has limited commitment power in the evaluation of bids. This assumption reflects the feature that program managers have short accountability spans because of frequent job rotations and turnovers, while a typical procurement cycle spans a long period.[6] The quality distortion disappears when the buyer myopia/limited-commitment assumption is dropped.

The rest of the paper is organized as follows. The model is introduced in Section 2. In Section 3, the effect of profit policy is illustrated in a simple model where firms can pick from only two different levels of quality. A more general model is presented in Section 4, where the Averch-Johnson effect is identified in a competitive environment. A numerical example is presented in Section 5, and Section 6 concludes.

---

[4]The intensity of nonprice competition is indirectly revealed by the fact that evaluation boards typically assign more than 50 percent of weight to nonprice factors (Fox, 1974).

[5]The underlying contractual situation where a procurer designs a mechanism of determining quality (or typically quantity) and transfer payment in an auctioning environment has been studied before in the context of optimal contract design (Laffont and Tirole, 1987, McAfee and McMillan, 1987, and Riordan and Sappington, 1987, to name a few). This paper is differentiated from this literature in its focus on the particular regulatory setting.

[6]The average tenure of program managers is two or three years, while a typical procurement spans a decade (Fox, 1988).

2. THE MODEL

Following Bower and Osband (Chapter Four), we consider a stylized procurement scenario, in which a competitively selected contractor comes under the regulatory regime of profit policy at some point of the procurement cycle. Transition to the regulatory regime can be triggered by either program revision or sole-source reprocurement.

Thus, the model has two periods: In the first period, a contractor is selected through two-dimensional bidding; in the second period, the firm receives profits on the basis of cost-based pricing. The significance of each period is weighted by λ and $1 - \lambda$, respectively. This weight reflects the length of each regime and discounting of the second period. Here we simply interpret λ as the relative length of the first period.

The first-period bidding is described as follows. A buyer, interpreted as DoD, solicits bids from $N(\geq 2)$ firms. Each bid specifies an offer of instantaneous quality, q, delivered for the length of the first period, and the lump-sum payment, p, for the first period. This instantaneous quality cannot be changed over the entire contract duration once it is chosen at the beginning.[7] The instantaneous quantity of output to be delivered is fixed and normalized to be one. In reality, the quality offer includes specification of technical characteristics, a delivery schedule, and other performance attributes. Our single-dimensional quality can be interpreted as a composite measure of these multidimensional quality attributes. Each proposal denoted as a pair $(q,p) \in \Re_+^2$ is then evaluated by a predetermined scoring rule $S(q,p)$. The firm that earns the highest score is declared the winner[8] and performs the offered contract terms for the first period.[9]

The buyer's first-period utility from a contract, (q,p), is given by:

$$U(q,p) = \lambda V(q) - p.$$

We make the following assumptions about the instantaneous surplus function.

Assumption 1: For all $q \in (0, \infty)$, $V(q)$ is bounded above from ∞, $V'(q) > 0$, $V''(q) < 0$, and $\lim_{q \to 0} V'(q) = \infty$, $\lim_{q \to \infty} V'(q) = 0$.

This assumption implies that the buyer's surplus increases with quality at a diminishing rate, and asymptotes to a finite value. The last two conditions are invoked to ensure an interior solution.

A firm i, upon winning, earns a profit in the first period from a contract (q,p):

---

[7] In the procurement of highly advanced weapon systems, major quality changes require costly investment. The findings of this paper still hold when this assumption is relaxed as long as there exists some degree of quality rigidity.

[8] If more than one firm achieves the highest score, one firm is randomly selected. Tie-breaking rules have no effect when the distribution has no atom.

[9] Using the terminologies of Che (1993), this rule is called a "first-score auction." There are other rules called "second-score" and "second-preferred-offer" auctions under which the winning firm is required to match the highest *rejected* score. See Che (1993) for the details.

$$\pi_i(q, p) = p - \lambda c_i q,$$

where c_i denotes firm i's instantaneous marginal cost of quality (henceforth simply called "marginal cost"). Losing firms earn reservation profits, normalized at zero. Before bidding, the marginal cost c_i is realized as private information of each firm i. The buyer knows only the distribution function of the cost parameter. The marginal cost c_i is independently and identically distributed over $[\underline{c}, \overline{c}]$ $(0 < \underline{c} < \overline{c} < \infty)$, according to a distribution function F for which there exists a positive, continuously differentiable density f. Because of complete symmetry among firms, the subscript i is suppressed hereafter for notational simplicity. Throughout the analysis, the following assumptions are made:

Assumption 2: $F(c)/f(c)$ is nondecreasing in c.

Assumption 3: The buyer never cancels procurement.

Assumption 2 is a standard regularity condition that often appears in the mechanism design literature. It holds if the density $f(\bullet)$ does not increase too fast, and is satisfied for standard distributions, including uniform, exponential, and the normal distribution. Assumption 3 ensures that the contract is awarded for all cost types.[10]

At the end of the first period, the firm's cost is audited. We assume that as a result of the auditing, the buyer obtains an unbiased estimate of the firm's cost.

In the second period, the firm negotiates a renewal contract with the DoD according to profit policy regulation. The instantaneous quality q, once determined in the first period, cannot be changed in the second period.[11] Thus, the same quality q (chosen by the firm in the first-period auction) is specified in the renewal contract. We also assume that the expected second-period cost is perfectly correlated with the first-period cost. (So the audited cost is also an unbiased estimate of the second-period cost.) Allowing partial correlation does not add any new insight.[12] The profit policy results in cost-based pricing that we model as a simple markup rule. The firm's revenue is determined as the markup $(1 + m)$ over the end-of-the-first-period cost estimate. Thus, the firm's expected second-period profit is: $(1 - \lambda)mcq$, and the buyer's expected utility is $(1 - \lambda)(V(q) - (1 + m)c\,q)$.

In summary, two-period expected profit for a winning firm is given by:

---

[10]Without the assumption, the solution of auction games will remain qualitatively the same but will involve a reserve score.

[11]In the defense procurement context, changing quality levels require costly research and development investment. Also, the quality does not have to be perfectly rigid over time. Partial rigidity will yield qualitatively the same result.

[12]The possibility that the second-period cost is reduced by the firm's effort from its first-period cost is considered by Bower and Osband (Chapter Four). When the effort is unobservable to the buyer, the resulting moral hazard problem leads the buyer to design the profit policy regulation in a similar way to the Averch-Johnson effect in this paper.

$$\pi(q,p \mid c) = p - \lambda c q + (1 - \lambda) mcq$$

$$= p - [\lambda - (1 - \lambda) m] c q, \tag{1}$$

when the firm has marginal cost c and offers (q,p) in the first-period bidding. Similarly, two-period expected utility for the buyer is given by:

$$U(q,p|c) = \lambda V(q) - p + (1 - \lambda)(V(q) - (1 + m)c q$$

$$= V(q) - [p + (1 - \lambda)(1 + m)c q]. \tag{2}$$

Note that the buyer's two-period payment is the sum of the fixed payment p in the first period and the cost-based payment $(1 - \lambda)(1 + m)c q$ in the second period.

Finally, a scoring rule must be specified. In principle, a scoring rule $S(q,p)$ can be any arbitrary function of (q,p) that the buyer may want to commit to. Here, we make two assumptions. First, the buyer cannot commit to a scoring rule differing from its true preference over (q,p). This assumption is sensible because it is practically very hard for the DoD to communicate its preference over complicated technical tradeoffs, especially in a way verifiable to a third party. The possibility of bid disputes also causes procurement officials to abstain from disclosing their evaluation procedures—a necessary condition for commitment to a different scoring rule. Second, we assume that the buyer is myopic: The scoring rule reflects the buyer's first-period preference. This assumption is again consistent with the special feature that program managers have short tours of duty and major procurement spans decades. From these two assumptions, it follows that the only feasible scoring rule is one that reflects the buyer's first-period preference:

$$S(q,p) = \lambda V(q) - p. \tag{3}$$

Before closing this section, a comment is warranted on the wisdom of this two-period procurement mechanism. Given that the winning firm's cost is revealed through auditing at the end of the first period, it is theoretically possible to design a more efficient mechanism contingent on end-of-the-first-period auditing. Although such a mechanism would be an interesting theoretical possibility to consider, it is not likely to be feasible; long contract duration typical in DoD procurement makes it difficult for an initial competitive contract to depend on the cost audits that are not immediately available. Furthermore, this kind of optimal contract approach is not consistent with the buyer's myopia and lack of commitment power that is assumed in most of this paper. As it will be shown in Section 4, our two-period institution can be made to approximately implement the first-best outcome when this assumption is relaxed. Another benefit from using this model is the clear comparison it permits with Bower and Osband (Chapter Four), which will help highlight the result obtained in this paper.

3. A SIMPLE ILLUSTRATION OF THE AVERCH-JOHNSON EFFECT

For a clear comparison with previous literature, it is useful to first consider the case where all firms' quality choices are fixed at some positive level, $\bar{q}$. With quality pre-fixed, the first-period bidding is reduced to a standard single-dimensional auction, and our two-period model becomes identical to that of Bower and Osband (Chapter Four) and closely resembles that of McAfee and McMillan (1987). Without going through a detailed review, I present a simple graphical illustration to highlight the intuition of the literature.

Because of the revenue equivalence theorem (Riley and Samuelson, 1981), the first-period bidding can be represented by a second-price auction. In the second-price auction, each firm bids its total cost of production described in Eq. (1). Figure 5.1 succinctly illustrates the equilibrium bidding outcome when the firms expect the profit policy regulation to be effective for the recontracting stage.

Suppose the buyer chooses zero markup $m = 0$. This means that the winning firm simply receives cost reimbursement in the second period. It follows from Eq. (1) that the two-period *effective* cost of production is $\lambda c \bar{q}$ for a firm with marginal cost c, and this is the equilibrium bid for the firm. The line $\overline{ab}$ in Figure 5.1 represents the equilibrium bid for each cost type of firm. Let c_2 be the realized second-lowest cost among all participating firms. Then, in equilibrium, the firm with the lowest marginal cost wins and pays $\lambda c_2 \bar{q}$. If the winning firm's marginal cost is c_1, the
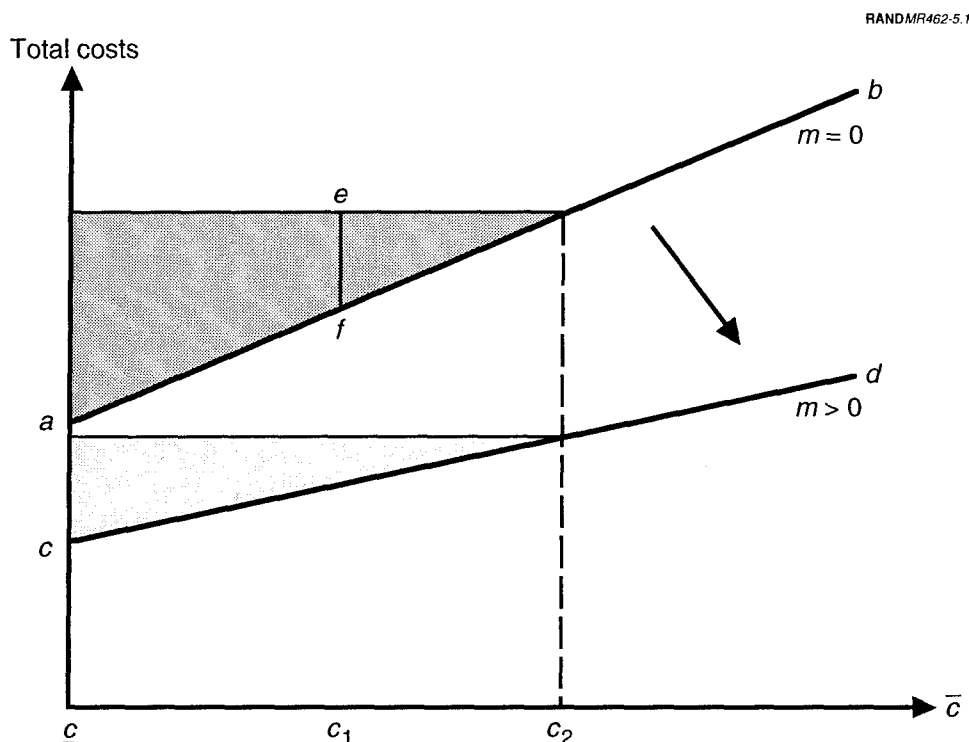

RAND*MR462-5.1*

Figure 5.1—Markups and Government Costs When Quality Is Given

difference between $\lambda c_2 \bar{q}$ and its own total cost becomes the profit (represented by $\overline{ef}$). Therefore, the dark-shaded area, when weighted by the probability of the winning firm's cost, represents the expected profits accruing to the winning firm conditional on the given realization of c_2. Roughly speaking, the objective of the buyer is to minimize the area for random c_2.

Now, suppose the buyer imposes a positive markup $m > 0$. Notice that this essentially creates a differential cost subsidy $(1 - \lambda) \, mc\bar{q}$ to each cost type c. Higher-cost types get more favorable subsidies, since the regulated profit $(1 - \lambda) \, mc\bar{q}$ is increasing in c. In Figure 5.1, this is represented by the fact that the new total cost curve $\overline{cd}$ of each firm is less steep than the original cost curve. As can be seen, the shaded area representing the information rents shrinks as a result of the positive markup. The intuition behind this is that the systematic "handicapping" of the low-cost type through the markup forces it to bid more aggressively, squeezing the profits accruing to the firm. In this simple model, actually, the first-best outcome can be approximated by rotating the cost curve all the way toward the horizontal axis; i.e., by raising the markup m toward $\lambda/(1-\lambda)$ (Proposition 1 of Bower and Osband, Chapter Four).

Now, let us introduce the quality choice in this model. As a simple version, we consider a case where there are only two technically feasible quality levels; i.e., $q \in \{q_i, q_h\}$ where $q_i < q_h$. This simplified version can be of some independent interest: Frequently, only a few design options are available for a particular weapon system. For instance, the two quality choices in this simple version can represent conventional and advanced design levels available to build a system. Che (1993) shows that the first-score auction yields the same performance as the second-score auction in which the winning firm is required to match the highest rejected score while choosing any quality level in its best interest in meeting the requirement. So, without any loss of generality, we consider the second-score auction.

Suppose first $m = 0$. Che (1993) shows that in the second-score auction, each firm picks quality $q \in \{q_i, q_h\}$ that maximizes $\lambda(V(q) - cq)$ and bids its two-period total cost, $(\lambda - (1 - \lambda)m)cq$ (see Eq. (1)). Let there be $\hat{c} \in (\underline{c}, \bar{c})$ that satisfies $V(q_i) - \hat{c} \, q_i = V(q_h) - \hat{c}q_h$. Then, in our two-quality context, this means that each firm bids the high quality if and only if its cost is less than the cutoff level $\hat{c}$. In Figure 5.2, $\overline{abcd}$ represents the equilibrium bid (= total cost) when each type picks quality in this way. The equilibrium bid jumps at the cutoff cost level, reflecting the higher costs associated with producing the high quality. Notice also that $\overline{ab}$ is steeper than $\overline{cd}$. Again, fix the second-lowest-cost type, c_2. That the same second-highest score applies to any winning bidder indicates that $\overline{hi} = \lambda(V(q_h) - V(q_i))$; i.e., a high-quality producer receives a premium over a low-quality producer equal to the utility increase. Also, the cutoff type $\hat{c}$ must be indifferent between producing q_i and q_h. That is, $\overline{hb} = \overline{ic}$. As before, the shaded area roughly measures the expected profits accruing to the winning firm conditional on the given realization of the second-lowest-cost type c_2.
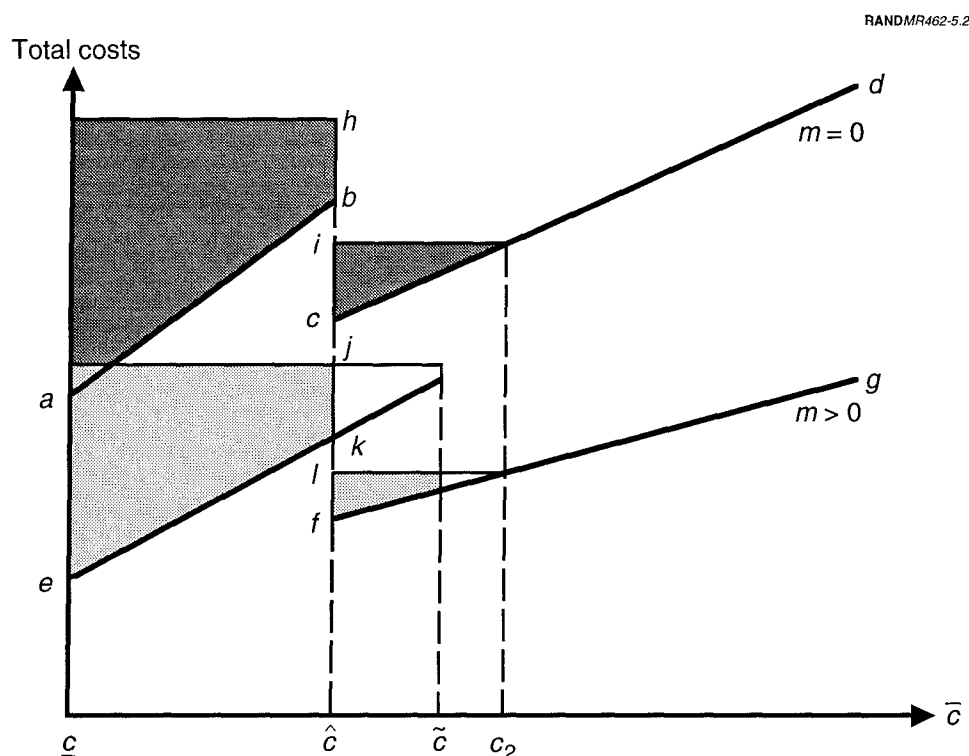
Figure 5.2—Markups and Government Costs When Quality Is Variable

Now, suppose the buyer chooses a positive markup. Then, as in the fixed quality case, each cost type faces a lower, less-steep cost curve ($\overline{ekfg}$) for producing each quality. As before, the expected profits are squeezed because of more aggressive bidding competition. However, the improved competition is not gained for free. The positive markup affects the equilibrium quality choice. Specifically, it makes the high quality a more favorable choice. To see this, consider the quality choice of the original cutoff type $\hat{c}$. The price premium for producing high quality is the same as before; i.e., $\overline{jl} = \overline{hi}$. But the effective cost of producing the high quality has become relatively lower than that of producing the low quality because the high quality entails a greater rate base. Thus, the profit margin for the original cutoff type is higher when it produces the high quality ($\overline{jk} > \overline{lf}$). This implies that the type $\hat{c}$ is no longer indifferent between q_i and q_h: It is strictly better off producing the high quality. In fact, every type from $\hat{c}$ up to the new cutoff point $\tilde{c}$ switches over to the high quality. This illustrates the Averch-Johnson effect in quality distortion: Firms are likely to offer a higher than optimal level of quality to increase the second-period rate base.

4. THE GENERAL MODEL OF PROFIT POLICY REGULATION

This section formally presents the idea illustrated in the previous section. Here, we consider a more general case where the winning firm can pick any quality $q \in [0, \infty]$.

The first part of the analysis involves identifying the equilibrium contract terms determined in the two-period contracting game. Since the second period entails no strategic play, we focus on the first-period auction game. To identify equilibrium contract terms in the first-period auction, we follow steps analogous to those adopted in Che (1993).

First we consider each firm's quality offer. Suppose a firm with marginal cost c considers offering a pair (q,p) that will result in a score, $\bar{s}$. For that target score $\bar{s}$, the firm chooses q to maximize the two-period profit $\pi(q, p \mid c)$ subject to the constraint $S(q,p) = \bar{s}$ Given the scoring rule $S(q,p) = \lambda V(q) - p$, the optimal quality offer for the firm with type c is

$$q_R(c) = \text{argmax } V(q) - \phi(m)cq,$$

where $\phi(m) \equiv 1 - [(1 - \lambda)/\lambda]m$. We first note that this quality choice does not depend on the target score level $\bar{s}$. (This property comes from additive separability of the scoring rule.) That is, this quality offer is a (weakly) dominant choice for the firm. Second and more important, the firm recognizes the two-period modified cost $\phi(m)c$ as an effective marginal cost. The modified cost is equal to the firm's true two-period cost c minus a term, $[(1 - \lambda)/\lambda]m$ This second term captures the beneficial effect of increasing quality on the firm's second-period rate base. Each firm, in choosing quality, effectively discounts this benefit as a cost. It also follows from this that the markup must be set at $m < \lambda/(1-\lambda)$, since otherwise, each firm will pick arbitrarily large quality. Without loss of generality, we restrict attention to $m < \lambda/(1-\lambda)$ from now on.

Once we obtain the dominant strategy quality offer, we can reduce the two-dimensional auction to a single-dimensional one by appropriately relabelling variables. Let $S_0(c) = \max V(q) - \phi(m)cq$ for all $c \in [\underline{c}, \overline{c}]$. Then, $S_0(\bullet)$ is strictly decreasing, and therefore its inverse exists. Now, consider the following change of variables:

$$v \equiv S_0(c), \qquad H(v) \equiv 1 - F(S_0^{-1}(v)), \qquad b \equiv S.$$

The problem can then be reinterpreted as one in which each firm, indexed according to its productive potential v with cumulative distribution $H(\bullet)$, proposes to meet the level of score, S. In particular, letting $b(\bullet)$ denote an equilibrium bid function of v that is symmetric and increasing, we can express each firm's expected profit in the first-score auction as:

$$\pi(q^*(c), p \mid c) = [p - \lambda\phi(m)cq_R(c)]Prob\{win \mid S(q_R(c), p)\}$$
$$= [v - b]H(b^{-1}(b)).$$

This is precisely the expected profit a bidder faces in a standard first-price auction. From the standard equilibrium result in this case (due to Riley and Samuelson (1981) among others), the following proposition is immediate.

Proposition 1: A unique symmetric equilibrium of a first-score auction is a pair $(q(\bullet), p(\bullet))$ for each firm that satisfies

$$q_R(c) = \arg\max \; V(q) - \phi(m)cq, \tag{4}$$

$$p_R(c) = \lambda\phi(m)cq_R(c) + \lambda\phi(m)\int_c^{\bar{c}} q_R(t)\left[\frac{1 - F(t)}{1 - F(c)}\right]^{N-1} dt, \tag{5}$$

for $m < \lambda/(1-\lambda)$.

Note that the second term of Eq. (5), which corresponds to a profit upon winning, is decreasing in m. As in the fixed-quality scenario, raising the markup forces firms to bid more aggressively and reduces the profits accruing to the winning firm. Also, if there is no markup ($m = 0$) and quality is fixed, the equilibrium price bid in Eq. (5) is reduced to that in a standard single-dimensional first-price auction.

Upon substituting the winning firm's equilibrium bid into the buyer's two-period utility in Eq. (2), we get

$$EU_R = E_{(1)}\left\{ V(q_R(c)) - \left[c + \lambda\phi(m)\frac{F(c)}{f(c)}\right]q_R(c)\right\}, \tag{6}$$

where $E_{(1)}[\bullet]$ denotes expectation over $\min\{c_1, c_2, \ldots, c_N\}$. The expression inside the square brackets is the so-called "virtual" marginal cost that the buyer incurs in inducing an additional unit of quality from the winning firm. It is the sum of the firm's physical marginal cost c and the extra term, often referred to as "informational cost" in the mechanism design literature. This extra term represents the cost associated with the buyer's inferior informational position regarding the realized cost of the winning firm. Observe that, other things being equal, m reduces the information cost and increases the expected utility of the buyer. In fact, one can eliminate the informational costs by raising the markup arbitrarily close to $\lambda/(1-\lambda)$. As m approaches $\lambda/(1-\lambda)$ from below, $\phi(m)$ goes to zero. When quality is fixed, this markup implements the first-best outcome (Bower and Osband, Chapter Four). The main point of this paper is that this markup strategy may not be optimal for the buyer if quality is variable.

To understand the welfare implication of the equilibrium quality $q_R(\bullet)$ in Eq. (4), consider an optimal quality schedule for the buyer. The optimal quality schedule for the buyer, denoted as $q_0(\bullet)$, can be obtained by directly performing point-wise maximization of Eq. (6) with respect to q_R. For a given markup m,

$$q_0(c) = \arg\max \; V(q) - \left[c + \lambda\phi(m)\frac{F(c)}{f(c)}\right]q. \tag{7}$$

Comparing the equilibrium quality in Eq. (4) with the optimal one in Eq. (7) reveals a possible quality bias. To measure this bias, consider the buyer's virtual marginal cost of inducing quality that each firm does not internalize in its equilibrium quality bid:

$$(1-\phi(m))c+\lambda\phi(m)\frac{F(c)}{f(c)}.$$

There are two sources of quality distortion. The first distortion comes from the fact that the firms do not internalize the buyer's informational costs, $\lambda\phi(m)F(c)/f(c)$, in their choice of quality offers. This effect always predicts that the winning firm will oversupply quality. The second distortion comes from the fact that $\phi(m)\neq 1$ if $m \neq 0$. This distortion can be interpreted as the Averch-Johnson effect. Notice that $\phi(m)<1$ if $m > 0$. A winning firm's first-period total cost is higher if it picks a higher-quality level in its bid. When the markup m is positive, this means that the firm's second-period regulated profit will be higher. Therefore, each firm, knowing this rate-base increasing effect, offers higher quality in the first-period bidding than otherwise. If $m > 0$, thus, the Averch-Johnson effect augments the first effect (henceforth called "informational effect"). If, on the contrary, the markup is negative, the same argument holds in the opposite direction. In this case, $\phi(m)>1$, and the Averch-Johnson effect actually causes each firm to lower quality in the bid. This effect can offset the quality bias created by the informational effect. In sum, we conclude that the quality bias is positive in equilibrium if the markup is positive or not too negative. That is, for any $c \in [\underline{c}, \overline{c}]$, there exists $\hat{m}< 0$ such that $q_R(c)> q_0(c)$ if and only if $m > \hat{m}$.

Several remarks can be made. First, the informational effect becomes negligible if the winning firm's marginal cost of quality is close to $\underline{c}$. Since as the number of bidders, N, tends to infinity, the winning firm's marginal cost becomes arbitrarily close to $\underline{c}$; this means that the first type of bias becomes less serious as the first-period bidding becomes more competitive. The same is not true for the Averch-Johnson bias. Clearly, this bias exists even for the most efficient type of firm as long as $\phi(m)\neq 1$. Hence, the Averch-Johnson effect does not disappear as competition intensifies. Second, λ has special implications on the two types of quality distortion. As the first-period contract becomes relatively more important than the second-period contract, the quality bias associated with the informational effect increases while the Averch-Johnson effect diminishes.

Now, we are in a position to determine the buyer's optimal choice of markup. In choosing the optimal markup, the buyer must recognize both the bidding competition effect (of reducing the winning firm's profit margin in Eq. (5)) and the quality bias effect associated with raising the markup. As mentioned above, when m is raised toward $\lambda/(1-\lambda)$, the winning firm's profit margin is reduced, but quality becomes excessive to the buyer because of both the informational and the Averch-

Johnson effects.[13] These two conflicting forces must be optimally weighed. In general, it is not easy to characterize the optimal markup. But the following proposition shows that the optimal markup m^* cannot be arbitrarily close to $\lambda/(1-\lambda)$.

Proposition 2: The optimal markup m^* is bounded above from $\lambda/(1-\lambda)$, and at the optimal markup the induced level of quality is excessive; i.e., $q_R(c) > q_0(c)$ for all $c \in [\underline{c}, \overline{c}]$.

Proof: To prove the first statement, suppose otherwise. Then, it must be optimal for the buyer to raise m arbitrarily close to $\lambda/(1-\lambda)$. Observe that as m approaches $\lambda/(1-\lambda)$ from below, $\phi(m)$ approaches from above to zero. From Eq. (4) and Assumption 1, then, $q_R(c)$ must tend to infinity for all $c \in [\underline{c}, \overline{c}]$. This, in turn, implies that the cost term in Eq. (6) goes to infinity while the gross surplus remains finite (because of Assumption 1). This, however, cannot be optimal for the buyer, since it has an option of cancelling the procurement and earns finite payoff instead of negative infinity. To prove the second statement, we differentiate the buyer's expected utility in Eq. (6) with respect to m:

$$\frac{dEU_R}{dm} = E_{(1)}\left\{(1-\lambda)\frac{F(c)}{f(c)}\, q_R(c)\right\} + E_{(1)}\left\{\left[V'(q_R(c)) - \left(c + \lambda\phi(m)\frac{F(c)}{f(c)}\right)\right]\frac{dq_R(c)}{dm}\right\}.$$

Suppose, contrary to the statement, at the optimal markup m^*, $q_R(c) \leq q_0(c)$ for $c \in [\underline{c}, \overline{c}]$ (it is not possible for this inequality to hold for only some subset of $[\underline{c}, \overline{c}]$). Then, at such m^*, $V' - (c + \lambda\phi F/f) \geq 0$, and $dEU_R/dm > 0$. This is a contradiction to the optimality of m^*.

This result stands in a sharp contrast to the main finding of Bower and Osband (Chapter Four). They argue that the optimal markup must be set sufficiently high so that each firm essentially faces zero cost after paying the first-period competitive bid; i.e., m must be arbitrarily close to $\lambda/(1-\lambda)$. This high markup approximates the first-best outcome by inducing firms to buy in very aggressively in the first-period bidding. This result does not hold if firms can freely choose quality in their bids. Because the quality bias becomes arbitrarily large as the markup is raised toward $\lambda/(1-\lambda)$, the optimal markup must be strictly below $\lambda/(1-\lambda)$. In fact, an optimal markup can even be negative if the welfare loss associated with quality bias is very large.[14] Finally, the second part of the proposition implies that the quality bias, often documented in the defense procurement literature, may be the result of the pro-

---

[13]This tradeoff becomes clear upon differentiating the buyer's expected utility Eq. (6) with respect to m:

$$\frac{dEU_R}{dm} = E_{(1)}\left\{(1-\lambda)\frac{F(c)}{f(c)}\, q_R(c)\right\} + E_{(1)}\left\{\left[V'(q_R(c)) - \left(c + \lambda\phi(m)\frac{F(c)}{f(c)}\right)\right]\frac{dq_R(c)}{dm}\right\}.$$

The first term captures the bidding competition effect and is positive for $m < \lambda/(1-\lambda)$. The second term shows the quality bias effect. It is negative for all $m \in [0, \lambda/(1-\lambda)]$, since Eq. (4) clearly shows that $V' - (c + \lambda\phi F/f) < 0$ for such m and that q_R is increasing in m.

[14]Bower and Osband (Chapter Four) point out a similar possibility when the winning firm makes a cost-reducing effort that the buyer cannot observe. In this situation, cost-based regulation creates a disincentive for the winning firm's cost-reduction effort, and the buyer, as in the variable quality situation, must balance conflicting effects associated with raising the markup.

curer's attempt to elicit bidding competition. This view is contrasted to that of Rogerson (1990) who explains the quality bias as a result of the procurement officers' inability to internalize a correct social objective function.

The critical assumption behind the above proposition is that the buyer is myopic and has limited commitment power in its design of the scoring rule. We show now that the quality distortion can be rectified if the buyer is able to commit to any scoring rule.

Let $q^* \equiv \text{argmax } V(q) - cq$. Consider the following scoring rule:

$$\hat{S}(q, p) \equiv V(q) - \left(p + (1 - \lambda)(1 + m) \int_k^q q^{*-1}(s)ds \right) \quad \text{if } q \in [q^*(\bar{c}), q^*(\underline{c})],$$

and $\hat{S}(q, p) = -\infty$, otherwise. Compared with the myopic scoring rule in Eq. (5), this new scoring rule has an extra penalty term that is increasing q. That is, other things being equal, $\hat{S}$ provides a smaller quality incentive than the naive scoring rule S. The following proposition shows that the first-best can be again approximately implemented in the variable quality context if the buyer uses this new scoring rule.

Proposition 3:

(i) The scoring rule $\hat{S}$ implements a quality schedule $q^*(\bullet)$ as long as $m < \lambda/(1 - \lambda)$.

(ii) With this scoring rule the buyer can approximately implement the first-best outcome by raising m arbitrarily close to $\lambda/(1 - \lambda)$.

Proof: To prove (i), following the logic of Proposition 1, it suffices to show that for any $c \in [\underline{c}, \bar{c}]$

$$q^*(c) = \text{argmax } \hat{S}(q, \phi(m)cq).$$

Taking the derivative of this with respect to q, we get

$$V'(q) - (1 - \lambda)(1 + m)q^{*-1}(q) - (\lambda - (1 - \lambda)m)c.$$

This becomes zero when $q = q^*(c)$. Thus, the first-order condition is satisfied at $q = q^*(c)$. Differentiating the above derivative with respect to q yields for $q \in [q^*(\bar{c}), q^*(\underline{c})]$

$$V''(q) - \frac{(1 - \lambda)(1 + m)}{q^{*'}(q^{*-1}(q))}$$

$$= V''(q)[1 - (1 - \lambda)(1 + m)]$$

$$< 0, \quad \text{if } m < \frac{\lambda}{1 - \lambda}.$$

Thus, the second-order condition is globally satisfied in the relevant region of q.

Next, we prove (ii). It follows from Eq. (7) that, as m is raised to $\lambda/(1-\lambda)$, $q_0(c)$ approaches q*(c) for all $c \in [\underline{c}, \overline{c}]$. Furthermore, the buyer's expected utility approaches $E_{(1)}\{V(q^*(c)) - cq^*(c)\}$, or its maximum attainable surplus level.

This result reconfirms the main proposition of Bower and Osband (Chapter Four) even in a variable quality setting. It identifies the buyer's myopia and its lack of commitment power as the ultimate source of the quality bias. Nevertheless, one must be cautioned that the mechanism featured in the proposition may not be easy to implement. First, committing to a complicated scoring rule like $\hat{S}$ may be practically impossible. Second, correcting buyer myopia often requires sweeping changes in employment arrangements as well as organizational cultures for the procurement employees. These kinds of changes are not just difficult to implement but may be undesirable from a broad perspective of the government's objective.

5. NUMERICAL EXERCISE

In this section, I present a numerical example of the model with two benefits in mind: (1) The numerical analysis highlights the practical importance of quality distortion—namely, the fact that even a small amount of distortion can produce a magnified cost impact; (2) it helps us understand the implications of the markup policy on interrelated phenomena such as quality distortions, costs of procurement, and cost overruns.

Consider the following numerical specification of the model:

(i) $V(q) = aq - 1/2\,bq^2$, where $a = 3$ and $b = .2$;

(ii) c is uniformly distributed over $[1,2]$;

(iii) $\lambda = .5$ and $N = 3$.

Table 5.1 shows expected quality distortions under different markup rates and their effects on the cost of procurement. Each distortion term is broken down into its two components: information cost effect and the Averch-Johnson effect. Furthermore, various stylized facts about defense procurement are tied together: Table 5.1 shows that a high markup induces large quality distortion, high total cost, and high cost overruns.

Observe that the total quality distortion (in percentage terms) is translated into a slightly magnified expected cost increase (also in percentage terms). As an interesting reference, Bower (1993) reports in his similar simulation that a cost saving from any possible improvement on the simple markup rule (into a more complicated menu, say) is about 1 percent (when the quality choice problem does not exist). When quality is endogenously chosen, Table 5.1 reveals that a typical cost saving

from a potential improvement on quality choice averages 10 or 12 percent![15] Although it is difficult to compare different simulation models, this stresses the significance of the role played by variable quality.

A breakdown of the quality distortion shows an interesting tradeoff between two types of quality distortion. As m increases, the information-cost-related quality bias decreases, confirming the claim that the increased bidding competition squeezes the information rents. On the other hand, the Averch-Johnson distortion grows with m. (Here the optimal m is near $-.2$.) Finally, the magnitude of the cost overrun increases with the markup. The magnitude is inversely related to the amount of buy-in in the first period. A high markup induces firms to buy in more aggressively, resulting in a higher cost overrun.

6. CONCLUSION

This paper has shown that the DoD's profit policy can create an Averch-Johnson type of quality bias, which can negate its potential benefit as a "handicapping device." The modeling framework adopted to show these results abstracts from some of the important features of quality determination in defense procurement. For instance, we assumed that bidding firms have perfect information about the technological frontier of producing quality before participating in the auction. In practice, however, firms are often still in the process of acquiring technologies when they make proposals. In this case, enforcing a particular design specification may be a problem,

Table 5.1

Quality Distortions and Markups

Markup (m)	−.2	−.1	0	.05	.1	.2
Quality	7.56	8.18	8.80	9.11	9.42	10.04
Bias: Information (%)[a]	9.53	8.07	6.82	6.26	5.74	4.78
Bias: Averch-Johnson (%)[b]	−16.40	−7.58	0.00	3.40	6.58	12.35
Total cost	10.10	10.90	11.69	12.07	12.45	13.19
Bias: Information (%)[c]	11.58	9.77	8.19	7.48	6.83	5.66
Bias: Averch-Johnson (%)[d]	−17.52	−8.06	0.00	3.59	6.92	12.84
Cost overrun (%)[e]	−43.20	−30.48	−15.22	−6.38	3.42	26.71

[a] $E[(q^* - q_0)/q_0] \times 100$; where q_0 is the level of optimal quality that maximizes Eq. (5) (on p. 92) and q^* is the first-best level of quality chosen under the naive scoring rule when $m = 0$.

[b] $E[(q_R - q^*)/q_0] \times 100$; where q_R is the quality level determined under the myopic scoring rule when the markup rate is m.

[c] $\{E[c(q^*) - c(q_0)]/E[c(q_0)]\} \times 100$; where $c(x)$ is total two-period cost when quality level x is produced.

[d] $\{E[c(q_R) - c(q^*)]/E[c(q_0)]\} \times 100$.

[e] $\{\{E[(1 + m)cq_R] - E[p_R(c)]\}/E[p_R(c)]\} \times 100$, which represents the rate of cost increase over the two periods.

---

[15]The figure represents the cost inflation associated with excessive quality choice caused by the quality bias. Note that this cost inflation exaggerates the efficiency loss since the government gets more quality with the increased cost.

and thus the procurer may attempt to devise a scheme, such as prototype competition, that will help increase the information about the feasibility of proposed design specifications. For a clear understanding of the procurement institution, future research addressing the special nature of quality will be needed.

INFORMATION EXPROPRIATION AND MORAL HAZARD IN OPTIMAL SECOND-SOURCE AUCTIONS

by Lars A. Stole[1]

1. INTRODUCTION

Amidst shrinking budgets and increasingly expensive technology, the U.S. government has begun to pursue more vigorously the use of competition to reduce the costs of its defense procurements. Witness the enactment of the Competition in Contracting Act of 1984, which reoriented the procurement process around competition and dramatically narrowed the use of sole-source procurement strategies. It is commonly recognized that this new emphasis has had a strong effect on increasing the use of competition in procurement, and in particular, second-sourcing.[2]

Several approaches exist for introducing competition into the acquisition process. When an accurate and complete description of the developing firm's technology (known as a data package) exists, the government may choose merely to advertise the procurement competition to interested and qualified bidders, offering a fixed-price contract to the lowest bidder. When an adequate data package does not exist, or when the item to be produced is very complex, the government may choose to make an "educational" contract with a second source; such a contract typically consists of the purchase of small quantities of the item (called "learning buys") at a higher relative cost, thereby providing the second source with production experience and absorbing some of its initial production setup costs.

Both of these methods require the transfer of data from the developer to a second source—a result that can be accomplished in various ways. The developer's contract may contain negotiated terms for license fees, or the firm may be required by law to turn over all data to the government for use in the competitive procurement. (In the

United States, a firm must turn over to the government all data when either public funds have entirely funded the development or the government has Government Purpose License Rights in the project's data.) This approach is known as directed licensing, whereby the developer licenses its technology to the second source.

Three policy questions of increasing complexity emerge concerning the licensing of technology to a second source. First, putting aside the developer's incentives for investment in cost-reducing technology, when (if ever) should the government introduce a second source rather than remain with the developer? Second, when the second source has the added option of using its own designs and technologies, rather than the developer's data package, an additional question arises: When should the government choose to have the developer license the second source rather than allow the second source to produce using its own technology? That is, how should the government approach the problem of choosing among the developer, licensing a second source, and selecting a second source but directing it to use its own technology. Third, turning to the problem of moral hazard, when the cost of the product depends in part on the unobservable investments of the developer during initial stages of product design, how should the government respond with its technology transfer policy? Ostensibly, second-sourcing may have the deleterious effect of reducing *ex ante* investment by the developer.

We begin by examining a simple model of second-sourcing in which a tradeoff exists between inefficiently transferring technology and reducing the profits of the firms (and hence reducing the price of procurement). It is possible that introducing such a cost inefficiency is optimal for the buyer if it can also reduce the rents that firms receive from their private information. As in Baron and Myerson (1982) and Laffont and Tirole (1986), the buyer's goal is to balance the introduction of inefficiencies with the reduction of these "information rents" so as to obtain the lowest possible expected price. Our approach differs from these previous works as we consider the role of competition via second-sourcing as the rent-reducing inefficiency.

In Sections 2 and 3, we consider the situation where the buyer can commit to a take-it-or-leave-it offer, but all contracts are constrained to be *ex post* profitable.[3] The model consists of one buyer (the government) and two sellers (a developer and a second source). The government has three procurement alternatives: Choose the developer to produce, choose the second source to produce using its own technology, or choose the second source to produce using technology transferred from the developer. When second sources do not have their own technology to produce, as is a common situation in defense procurement, the government's options are more restricted, but the analysis is easily incorporated below. At the contracting stage, all parties have symmetric information, and the government commits to specific rules for an auction it will later conduct. The sellers determine their costs and bid accordingly; the rules establish who will produce and how much each seller will receive as a function of the bids.

---

[3]This *ex post* individual rationality constraint is also known as limited liability in the contracts literature. See Sappington (1983).

A rule that optimally transfers technology induces both the developer and the second source to report their costs truthfully while leaving each with less rents from their private information. Intuitively, the existence of a second source allows the buyer to compete away some of these information rents via an auction, where the licensing option can be thought of as the addition of a third seller. Although this additional bidder may have higher costs, it also has less of an informational stake in the transferred technology: If the production costs using the transferred technology are less related to the second source's own costs and more related to the developer's costs, its requisite information rent for truth-telling will be significantly lowered. Consequently, informational rent-reducing gains from technology transfer exist.[4]

Although a policy of transferring information-laden technology may reduce rents, we might suspect that such a policy would have perverse effects upon the developer's initial incentive to invest. In Section 4, this paper endogenizes the developer's investment decision and derives the optimal auction in the moral hazard environment. The results indicate that the solution to the moral hazard problem entails a change in the probability of choosing production by the developer as a function of the project's reported production cost using the developer's technology. The probability of licensing is more sensitive to the developer's announced cost when moral hazard considerations are present.

This paper has important policy ramifications. First, with respect to our earlier policy questions regarding the optimal transfer rule, we find that a commitment to transferring technology for some bids may reduce expected procurement costs, even when moral hazard is present. Second, this paper provides a caveat for the common empirical practice of evaluating the gains from licensing by comparing the posttransfer cost of production with the estimated cost of production by the developer. Such a comparison ignores the *ex ante* gains in reduced information rents that result from the government's commitment to breakout technology for *bad* bids and it ignores the costs of reduced incentives for initial development.

The contributions of this paper, however, are not restricted to defense procurement. On a more general level, this paper considers the transfer of information-inherent "property" from one agent to another so as to reduce information rents. Such a strategy achieves rent reductions by *expropriating the agent's hidden information—* transferring property in which the information is embodied to a competing agent with a lower informational stake in the property. Providing that an alternative agent can use the asset, an optimal transfer has the potential for reducing the principal's acquisition costs.

2. THE MODEL

We present a model of a risk-neutral buyer with full-commitment ability and two risk-neutral sellers who are subject to limited liability. For exposition, we initially

---

[4]The idea of transferring the information-inherent component of one agent to another so as to reduce information rents is not entirely new to the literature. Riordan and Sappington (1989), for example, make use of such transfers in their examination of defense procurement second-sourcing.

consider the problem of a buyer (the government) who must procure an item from one of two potential sellers (firms). The government desires to procure a single object at the lowest possible cost. It proposes a take-it-or-leave-it contract to both sellers: the primary seller (firm 1) and the secondary seller (firm 2). The contract commits the buyer to deal with the sellers in a prespecified manner after the sellers have announced their costs, and must guarantee both sellers nonnegative income. Each firm either accepts or rejects the contract. Following their decision, they discover their production costs. The government does not observe costs either *ex ante* or *ex post*.[5] After learning their costs, firms make announcements (i.e., "bids") to the government, who chooses which firm will produce; in the case that the second source is chosen, the government additionally chooses whether or not to transfer the developer's technology. We assume that the government's valuation is sufficiently large that it always chooses to procure the item. Monetary payments are made in accordance with the initial contract.

We may think of the contract that the buyer offers as a commitment to use a specific auction mechanism. In this way, we analyze the problem of choosing the optimal contract as one of optimal auction design. In particular, we will consider truthful revelation mechanisms, using techniques similar to those found in Myerson (1981).

Each firm's cost, c_i, is independently distributed according to the continuous probability density, $f_i(c_i) > 0$, on a compact set, which we take to be [0,1] without loss of generality. $F_i(c_i)$ is the corresponding cumulative distribution function, and we make the common regularity assumption that $F_i(c_i) / f_i(c_i)$ is nondecreasing in c_i.[6]

The total cost to firm 2 of producing with firm 1's technology, i.e., the total cost of production under licensing, is given by the function $\ell(c_1, c_2)$, which includes the cost of transfer, if any. We will further assume that $\partial\ell(c_1, c_2) / \partial c_2 = \ell_2$, a constant; that is, $\ell(c_1, c_2)$ is linear in c_2. This implies that the second source's marginal cost effect on the licensed production is independent of the developer's cost, allowing us to separate the information effects from each other. We also assume that $1 > \ell_2 \geq 0$ and $\partial\ell(c_1, c_2) / \partial c_1 < 1$. Consequently, the second source has less informational stake in the transferred technology than its own technology.

The cost distributions and ℓ are common knowledge to the buyer and the sellers. We will often consider a particular situation with linear licensing costs.

$$\ell(c_1, c_2) \equiv \lambda c_1 + (1-\lambda)c_2 + \gamma, \tag{1}$$

[5] Laffont and Tirole (1986) and McAfee and McMillan (1986) consider contexts where the government can observe costs *ex post* but is unable to observe the firms' effort levels. The approach taken here differs from theirs because cost remains unobservable, but similar gains from technology transfers could be realized under alternative models with contractible costs.

[6] This regularity assumption is commonly referred to as the monotone likelihood ratio property (MLRP). Among others, the uniform, normal, logistic, chi-squared, exponential, and Laplace distributions satisfy this property.

for $1 \geq \lambda > 0$. With linear licensing costs, a proportion λ of the technology is transferable to firm 2 for a fixed transfer cost γ. In the extreme case of perfectly and costlessly transferable technology we have $\ell(c_1, c_2) = c_1$.

Finally, using cost reports, the government chooses from one of three possible production alternatives: (i) primary production; (ii) secondary production; and (iii), technology transfer or licensing (i.e., secondary production with technology transfer). For tractability, we do not include the logical fourth possibility of transferring technology from the secondary firm to the primary firm. It is important to note that in this framework, seller 2 can be required to produce using seller 1's technology, even when it is inferior to seller 2's own technology. In the defense procurement context, this assumption is plausible as technologies are easily verified. When second sources do not have their own designs to produce, the government's options are restricted to (i) and (iii) above, but the analysis incorporates this case by assuming that c_2 is sufficiently large and ℓ_2 is sufficiently small.

Along with the production decision, the government determines payments to each firm based upon their cost reports. A crucial constraint is that the government must guarantee nonnegative profits for both producers for all possible realizations of cost: No policy can be enforced *ex post* that would unduly harm a truthful seller. Here, we assume that no firm can be forced to accept a loss, which prevents the government from effectively buying the project from the sellers for the expected minimum cost of the production among them. By law, corporate bodies are protected from liability beyond the value of their assets. Our assumption is stronger but justified for several reasons. First, the assumption approximates a firm that is extremely risk averse beyond a certain level of losses. Given that managers are sensitive to excessive losses, it is plausible that the firms' behavior may be risk neutral over a moderate range but risk averse for dramatic losses. Additionally, from a purely descriptive perspective, it is doubtful that the government could force a defense company to continue production when it suffers excessively large losses. Boards of Contracts Appeal (BCAs), the neutral tribunals that have jurisdiction over government contract disputes, frequently grant equitable adjustments to contracts that impose excessive sacrifices upon firms. To this extent, a limit exists to the losses that a contractor can be forced to bear.

We do not allow the government's payment to the primary firm to depend upon any *ex post* discoveries made by the licensed firm after a transfer. If the buyer could do this, the first-best solution would be approximated by employing the secondary firm with an arbitrarily small probability to check the truth-telling of the developer, and then punishing this firm sufficiently hard whenever untruthful reports occur. This paper considers the more subtle issue involved when payments cannot be conditioned on an *ex post* report of another agent. Such a restriction appears realistic in the defense procurement context; otherwise, we would have to allow a time delay (perhaps years) between the auction and the agent's action (e.g., defense production) before enough verifiable evidence could be marshalled to levy a punishment against the primary agent.

3. THE OPTIMAL CONTRACT

In this model, the buyer commits to deal with the sellers in a predetermined manner after learning their reported costs. Using these reports, the buyer determines who produces the object, whether technology is transferred, and how much each firm shall be paid. The Revelation Principle states that without loss of generality, we may restrict our attention to direct revelation mechanisms. The class of mechanisms that we will consider is given by $M = \{\{\phi_i(c_1, c_2)\}_{i=0}^{i=2}\{t_i(c_1, c_2)\}_{j=1}^{j=2}\}$, where, for given reported costs, ϕ_i is the probability that production alternative i is chosen by the buyer, and t_j is the transfer to firm j. The production alternatives, $i = 0, 1, 2$, correspond to licensed production, firm 1 (developer) production, and firm 2 (second source with own technology) production, respectively.

The First Best

Before examining the optimal contract under limited liability and asymmetric information, we note the properties of the full-information contract. Under the full-information contract

(i) the most efficient form of production is chosen:

$$\phi_0(c_1, c_2) = \begin{cases} 1 & \text{if } \ell(c_1, c_2) \le \min\{c_1, c_2\}, \\ 0 & \text{otherwise,} \end{cases}$$

$$\phi_1(c_1, c_2) = \begin{cases} 1 & \text{if } c_1 < \min\{c_2, \ell(c_1, c_2)\}, \\ 0 & \text{otherwise,} \end{cases}$$

$$\phi_2(c_1, c_2) = \begin{cases} 1 & \text{if } c_2 < \min\{c_1, \ell(c_1, c_2)\}, \\ 0 & \text{otherwise;} \end{cases}$$

(ii) the buyer pays the producer realized cost:

$$t_1(c_1, c_2) = \begin{cases} c_1 & \text{if } \phi_1(c_1, c_2) = 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$t_2(c_1, c_2) = \begin{cases} c_2 & \text{if } \phi_2(c_1, c_2) = 1 \\ \ell(c_1, c_2) & \text{if } \phi_0(c_1, c_2) = 1 \\ 0 & \text{otherwise;} \end{cases}$$

(iii) the firms make zero profit.

Because there is full information, the limited-liability constraint is not binding, as zero profits may be guaranteed for all outcomes. The firms will be willing to accept the above contract, and the buyer obtains the object at minimum (in this case, actual) cost. Any contract yielding a lower expected price must necessarily violate in-

dividual rationality. Note that if $\ell(c_1, c_2) > \min\{c_1, c_2\}$ for all c_1, c_2, then licensing is never optimal under full information. We will see below that even when licensing would never be optimal under full information, licensing may be a desirable strategy by the buyer in environments of asymmetric information.

Asymmetric Information and Limited Liability

Under the assumption that the other firm is truthful, payoffs to each firm as a function of reported and true costs are

$$\pi_1(\hat{c}_1, c_2 \mid c_1) \equiv t_1(\hat{c}_1, c_2) - \phi_1(\hat{c}_1, c_2)c_1,$$
$$\pi_2(c_1, \hat{c}_2 \mid c_2) \equiv t_2(c_1, \hat{c}_2) - \phi_2(c_1, \hat{c}_2)c_2 - \phi_0(c_1, \hat{c}_2)\ell(c_1, \hat{c}_2),$$

where $\hat{}$ denotes the reported type. Because neither firm knows the other's cost when it must make its report, it is useful to consider the expected payoffs for each firm:

$$\pi(\hat{c}_1 \mid c_1) = \int_0^1 [t_1(\hat{c}_1, c_2) - \phi_1(\hat{c}_1, c_2)c_1] dF_2(c_2), \tag{2}$$

$$\pi(\hat{c}_2 \mid c_2) = \int_0^1 [t_2(c_1, \hat{c}_2) - \phi_2(c_1, \hat{c}_2)c_2 - \phi_0(c_1, \hat{c}_2)\ell(c_1, c_2)] dF_1(c_1). \tag{3}$$

The mechanism-design problem facing the buyer is given below as program P1:

$$\min_M \int_0^1 \int_0^1 [t_1(c_1, c_2) + t_2(c_1, c_2)] dF_1(c_1) dF_2(c_2) \tag{4}$$

subject to

$$\pi_j(c_j \mid c_j) \geq \pi_j(\hat{c}_j \mid c_j), \forall c_j, \hat{c}_j, \tag{5}$$

$$\pi_j(c_1, c_2 \mid c_j) \geq 0, \forall c_1, c_2. \tag{6}$$

The objective function is the expected value of the payments paid by the government for the procurement. This is minimized subject to constraints in Eqs. (5) and (6). Constraints in Eq. (5) ensure Bayesian truth-telling. Constraints in Eq. (6) represent the limited-liability constraints for all states of nature; note that this is not an expectation over payments, but actual payment.

Following Mirrlees (1971), Myerson (1981), and others, we simplify the truth-telling and limited-liability constraints and incorporate them into the objective function to ascertain the nature of the optimal auction to obtain our first result. All results are proved in the appendix.

Proposition 1: The set of $\{\phi_i(c_1,c_2)\}_{i=0}^{i=2}$, which solve P1, is the same as that which solves program P2 below using point-wise minimization over the space of probability distributions on the production alternatives:

$$\min_{\phi_0,\phi_1,\phi_2} \left\{ \phi_1 \left[c_1 + \frac{F_1(c_1)}{f_1(c_1)} \right] + \phi_2 \left[c_2 + \frac{F_2(c_2)}{f_2(c_2)} \right] + \phi_0 \left[\ell(c_1,c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right] \right\}. \tag{7}$$

Optimal payments, which correspond to the solution of P2, are given by

$$t_1(c_1,c_2) = \int_{c_1}^{1} \phi_1(c_1,c_2)dc_1 + \phi_1(c_1,c_2)c_1, \tag{8}$$

$$t_2(c_1,c_2) = \int_{c_2}^{1} [\phi_0(c_1,c_2)\ell_2 + \phi_2(c_1,c_2)]dc_2 + \phi_2(c_1,c_2)c_2 + \phi_0(c_1,c_2)\ell(c_1,c_2). \tag{9}$$

Note that when $\ell(c_1,c_2) = c_2$, the proposition reduces to the standard auction result, which may involve handicapping if the cost distributions differ, such as in Myerson (1981). When firm i is chosen to produce, the transfers given in Eqs. (8) and (9) indicate that the losing firm receives nothing (this is because $\phi_{-i}(c_1,c_2) = 0$ and the integrand in the losing bidder's transfer function is zero given the monotonicity of the optimal choice functions). The transfer to the winner covers the costs of production and an additional rent term, which indirectly depends upon the report of the loser through the effect of the report on the integrand. The government could alternatively pay each bidder its expected information rents, thereby removing this interdependence of payments on reports, but under such a payment scheme the loser would typically receive some rents as well as the winner.

To understand the mechanics of this solution to the optimal auction, define the following variables as the virtual costs of each production alternative:

$$J_i(c_1,c_2) \equiv c_i + \frac{F_i(c_i)}{f(c_i)}, \quad i=1,2$$

$$J_0(c_1,c_2) \equiv \ell(c_1,c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)}.$$

Thus, the solution to P2 amounts to selecting the alternative with the minimum virtual cost. It will also be useful for a graphical analysis to define the following state-space partition over the set of all possible realizations of cost, where Ω^i is the set of (c_1,c_2) such that alternative i has the lowest virtual cost. That is, $\Omega^i \equiv \{(c_1,c_2) \mid J_i(c_1,c_2) \leq \min_k J_k(c_1,c_2)\}$. The following corollary flows directly from the definitions and the optimization of P2 in Proposition 1.

Corollary 1: The optimal auction consists of setting $\phi_i(c_1,c_2) = 1$ *iff* $(c_1,c_2) \in \Omega^i$.

The sets Ω^0, Ω^1, and Ω^2 represent cost realizations where licensing, developer production, and second-source production are chosen, respectively. Note that it is never strongly optimal to randomize between alternatives. The payments that implement the choices in P2 are determined using standard techniques. In all but the worst states, the above payment scheme pays positive rents to the firm chosen to produce, while the other firm receives nothing.

The Value of Technology Transfers

The commitment to use technology transfers under some cost realizations reduces *ex ante* information rents by relaxing firm 2's incentive compatibility constraints. Firm 2 can "less easily" say that it has high costs, because the buyer can always transfer firm 1's technology for it to produce.

To understand the intuition behind the optimal auction, consider the following polar case: $\ell(c_1, c_2) = c_1$. That is, firm 1's technology is completely and costlessly transferred under licensing to firm 2. For symmetry in this case, also assume that technology can be transferred from firm 2 to firm 1, completely and costlessly. Now a buyer may offer the following contract to extract fully the rent: If $c_1 \le c_2$, transfer firm 1's technology to firm 2 and have firm 2 produce the project using firm 1's technology for payment c_1; if $c_1 > c_2$, vice versa. Under this scheme, neither firm has an incentive to lie and the buyer completely extracts the information rents. Moreover, this scheme does not require firms to know each other's cost at the time of bidding.

Returning to our one-way technology transfer environment, transfers of technology under the optimal contract are *ex ante* optimal whenever $(c_1, c_2) \in \Omega^0$. An interesting question regards the determination of this region. Essentially, the buyer trades off the costs of inefficient licensing against the gain in reduced information rents. This is easily seen in the following proposition.

Proposition 2: The *ex ante* expected gain to the buyer from a policy of optimal licensing is given by

$$\int_{\Omega_1^0} \left(\frac{F_1(c_1)}{f_1(c_1)} - \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right) dF_1(c_1) dF_2(c_2) + \int_{\Omega_2^0} (1 - \ell_2) \frac{F_2(c_2)}{f_2(c_2)} dF_1(c_1) dF_2(c_2)$$

$$- \int_{\Omega_1^0} (\ell(c_1, c_2) - c_1) dF_1(c_1) dF_2(c_2) - \int_{\Omega_2^0} (\ell(c_1, c_2) - c_2) dF_1(c_1) dF_2(c_2),$$

where Ω_i^0 is the licensing region where alternative i would have been chosen if licensing were not available; i.e., $\Omega_i^0 \equiv \{(c_1, c_2) \mid J_0(c_1, c_2) < J_i(c_1, c_2) \le J_{-i}(c_1, c_2)\}$.

The proposition identifies two effects. The first two terms represent the gain to the buyer from information rent reductions. The last two terms represent the cost inefficiencies to the buyer from deciding on an inefficient production technique. The optimal contract can be reformulated as one in which Ω^0 maximizes the sum of the

terms. If no Ω^0 exists such that the sum is positive, the optimal contract does not entail licensing for any realization of costs. This suggests a corollary.

Corollary 2:   If $\ell(c_1, c_2) = \lambda c_1 + (1-\lambda)c_2 + \gamma, \lambda \in (0,1)$, and the sellers' cost distributions are symmetric on $[\underline{c}, \overline{c}]$, then an optimal auction will transfer technology with positive probability if

$$\lambda > \gamma f(\overline{c}).$$

If costs are distributed uniformly on $[\underline{c}, \overline{c}]$, then the optimal contract will utilize transfers if $\lambda(\overline{c} - \underline{c}) > \gamma$.

This result is in contrast to the result in Riordan and Sappington (1988) who find in their model without limited-liability constraints and without commitment that second-sourcing is rarely optimal. Because Riordan and Sappington do not assume limited liability, the firms compete away expected information rents at the initial symmetric information stage, so there is no information-rent problem. The gains from technology transfer in their model do not derive from reductions in information rents but from production enhancement: The government introduces less distortion in its decision of whether to produce at all if a second source exists as an alternative. This latter effect is absent in the present model because we have assumed for tractability that the government always procures the object—otherwise, we would find an additional positive term in Proposition 2, providing another gain to technology transfers.

An Example

Consider the following linear cost model with uniform distributions on $[0,1]$. That is, let $F_i(c_i) = c_i$, $i = 1, 2$, and let $\ell(c_1, c_2) = \lambda c_1 + (1-\lambda)c_2 + \gamma$. Thus, the virtual costs are given by $J_0(c_1, c_2) = \lambda c_1 + 2(1-\lambda)c_2 + \gamma$ and $J_i(c_1, c_2) = 2c_i$, $i = 1, 2$. For the initial case, we make the further simplifying assumptions that $\lambda = 1/2$ and $\gamma = 0$. The optimal partition over $[0,1]$ is graphed in Figure 6.1a as the projection of the minimum virtual cost onto the cost space.

The diagram indicates that when cost reports are relatively close, licensing is chosen. Intuitively, if the cost reports are relatively close, the licensing cost does not differ significantly from either the developer or second-source production, so there is little cost inefficiency from licensing. If firm 2 has a relatively low cost, it is expensive for the buyer to make the second source use the inefficient licensed technology rather than its own. Similarly, if firm 1 has a relatively low cost, it is productively inefficient to license technology to firm 2, since firm 1 is a superior producer. As costs become close, the losses in production inefficiencies shrink to zero and are offset by the gains from reduced information rents.

Ω_0 = License
Ω_1 = Developer production
Ω_2 = Second-source production

$$\lambda_a = \tfrac{1}{2}, \gamma = 0.$$

a

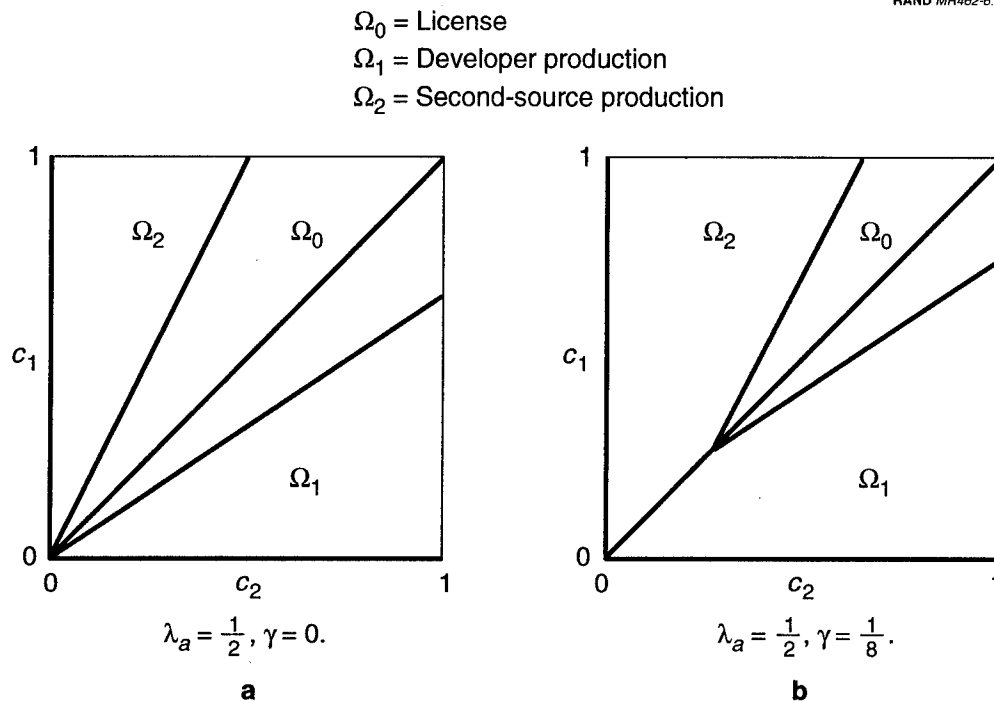$$\lambda_a = \tfrac{1}{2}, \gamma = \tfrac{1}{8}.$$

b

Figure 6.1—Optimal Auctions

We would expect the introduction of a fixed cost for transfer to increase the productive inefficiencies associated with licensing, and consequently the state space associated with licensing to shrink. To see the effect of a transfer cost, consider fixed licensing costs of $\gamma = 1/8$ as in Figure 6.1b. The licensing region has decreased substantially. As Corollary 2 predicts, if $\ell(c_1, c_2) = \lambda c_1 + (1-\lambda)c_2 + \gamma$ and costs are symmetrically and uniformly distributed on $[0,1]$, then there is no gain to licensing when $\gamma \geq \lambda$. As γ increases to $\lambda = 1/2$, the optimal licensing area shrinks to zero. More generally, Proposition 2 indicates that an increase in $\ell(c_1, c_2)$ (holding ℓ_2, c_1, and c_2 fixed) will reduce the probability of licensing and, if the increase is sufficiently large, will eliminate its use altogether. Mathematically, the costs of licensing (the latter terms in Proposition 2) increase while the benefits (the former terms) remain unchanged.

4. MORAL HAZARD

We naturally expect that in some situations where the initial agent (the developer) must make unobservable investments in reducing the marginal cost, c_1, of the final product, a policy of expropriating information via technology transfer would induce significant moral hazard. If the buyer can freely transfer the design to a second source to produce, the primary agent may have less incentive to reduce the marginal cost of production.

The Problem of Moral Hazard

This section extends the previous analysis by incorporating moral hazard on the part of the primary agent. We model this extension by assuming that the primary agent (the developer) may make cost-reducing investments. The question we ask is whether the buyer will find it optimal to favor the developer for cost-reducing investments in the award of the production contract, and if so, how?

As before, the approach we take is one of full commitment by the buyer and limited-liability constraints for the sellers. Initially, the buyer proposes a contract to the two sellers, which is accepted if it guarantees each nonnegative profit. Following the offer the developer chooses cost-reducing investment, e. This investment stochastically shifts (in a first-order sense) the distribution of the developer's production costs, c_1, and thereby improves the licensed cost of production as well. After investments have been made, costs of production are drawn by each firm from known distributions, with each firm's actual cost being observed only by that individual firm. The sellers then report their costs to the government. The government follows the agreed-upon contract and awards the production decision and payments conditional on the project's valuation and the cost announcements.

The resulting optimal contract is found to be a variation of the classical optimal auction design, which awards production to the most favorable virtual type. Under moral hazard, we find that the developer's virtual type has an additional term which decreases in production cost in a manner closely akin to the sharing rule in Holmström (1979). This suggests that in the stochastic cost-investment model, we would expect a discriminating auction to be used which may additionally favor the developer depending upon the resulting cost realizations.

The Model with Moral Hazard

The cost to firm 2 of producing with firm 1's technology is as before. The cumulative distribution function for the developer's cost is now given by $F_1(c_1 \mid e)$, and it is assumed that effort leads to a first-order stochastic improvement in the distribution on costs. For tractability, we will assume that $F_1(c_1 \mid e)$ satisfies the Concave Distribution Function Condition (CDFC), $\partial^2 F_1(c_1 \mid e) / \partial e^2 \leq 0$, which ensures us that the first-order approach to the principal-agent problem is valid.[7]

The cost to the developer for value-enhancing effort is given by $\psi(e)$, where $\psi(e)$ is increasing, strictly convex, $\psi'''(e) > 0$, $\psi(0) = \psi'(0) = \psi''(0) = 0$, and $\psi(1) = \infty$.

For simplicity in analyzing the moral hazard case, we assume that the government chooses from one of *two* possible production alternatives: (i) licensed production; or

---

[7]In addition to CDFC, a monotone likelihood ratio condition is usually required in pure moral hazard settings to assure that the agent's payoffs are monotonic in outcome. See Grossman and Hart (1983) and Rogerson (1985) for proofs of this proposition. (We use *concavity* in distributions rather than *convexity* as in Grossman and Hart, because higher costs are considered undesirable in our model.) With adverse selection, incentive compatibility requires that $\pi(c_1)$ be nonincreasing, and so we do not need an additional MLRP condition for sufficiency in the first-order approach. We may, however, have to solve the buyer's program subject to monotonicity of π in costs.

(ii) developer production; for simplicity, we ignore production by the second source using its own technology. Along with the production decision, the government determines payments to each firm based upon their cost reports. Again the crucial constraint is that the government must guarantee nonnegative profits for both producers for all possible realizations of cost.

The Optimal Contract Under Moral Hazard

The class of mechanisms considered is given by $M' = \{\{\phi_i(c_1, c_2)\}_{i=0}^{i=1}, \{t_j(c_1, c_2)\}_{j=1}^{j=2}\}$, analogous to before. The production alternatives, $i = 0,1$, correspond to licensed production and developer production, respectively.

The Choice of Investment. Consider first the investment decision. Given the assumptions regarding the distribution of costs, the developer's choice of effort solves

$$\max_{e \in [0,1]} \int_0^1 \int_0^1 \pi_1(c_1, c_2) dF(c_2) dF_1(c_1 \mid e) - \psi(e).$$

We can more simply characterize the solution to this program in the following lemma.

Lemma 1: A necessary and sufficient condition for the agent's optimal effort decision is

$$\int_0^1 \int_0^1 \left\{ \phi_i(c_1, c_2) \frac{\partial F_1(c_1 \mid e / \partial e)}{f_1(c_1 \mid e)} \right\} dF_1(c_1 \mid e) dF_2(c_2) = \psi'(e). \tag{10}$$

General Solution to the Contracting Problem. Having characterized the effort chosen by the developer for a given contract, we compute the buyer's optimal contract in the presence of moral hazard. To do so we simply append to the buyer's problem the additional condition from Lemma 1 to endogenize the investment decision. Call this program P3, and let μ represent the Lagrange multiplier associated with the investment constraint. Proposition 3 below provides the equivalence of P3 with a simple point-wise minimization problem.

Proposition 3: Assume that $F_{1,e}(c_1 \mid e) / f_1(c_1 \mid e)$ is nondecreasing in c_1, $F_1(c_1 \mid e) / f_1(c_1 \mid e)$ is nonincreasing in e, and $F_{1,ee}(c_1 \mid 0) = \partial[F_1(c_1 \mid 0) / f_1(c_1 \mid 0)] / \partial e = 0$. The set of $\{\phi_i(c_1, c_2)\}_{i=0}^{i=1}$, which solves P3, is the same as that which solves

$$\min_{\phi_i} \left\{ \phi_1 \left[c_1 + \frac{F_1(c_1 \mid \hat{e})}{f_1(c_1 \mid \hat{e})} + \mu \frac{F_{1,e}(c_1 \mid \hat{e})}{f_1(c_1 \mid \hat{e})} \right] + \phi_0 \left[\ell(c_1, c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right] \right\}$$

using point-wise maximization, where $\hat{e}$ is the buyer's expectation of firm 1's effort (which is correct in equilibrium) and $\mu > 0$. The level of effort, $\hat{e}$, induced by the buyer satisfies Eq. (10), and a set of optimal transfers are given by

$$t_1(c_1, c_2) = \int_{c_1}^{1} \phi_1(c_1, c_2)dF_2(c_2) + \phi_1(c_1, c_2)c_1 + \psi(\hat{e}),$$

$$t_2(c_1, c_2) = \int_{c_2}^{1} \phi_0(c_1, c_2)\ell_2 dF_1(c_1 \mid \hat{e}) + \phi_0(c_1, c_2)\ell(c_1, c_2).$$

The assumptions for the proposition regarding the monotonicity of $F_{1,e}/f_1$ in c_1 and F_1/f_1 in e are satisfied if e has more effect on reducing higher cost levels and the developer cannot increase information rents (i.e., the inverse hazard rate) by increasing investment.

The solution to the principal's problem has the same nature as the optimal auction without moral hazard, except that the state-space partition over firm production has been changed in an important way—it now depends more importantly upon the realization of the developer's cost. Consider the developer's virtual cost to the buyer:

$$\tilde{J}_1(c_1, c_2) \equiv c_1 + \frac{F_1(c_1 \mid e)}{f_1(c_1 \mid e)} + \mu \frac{F_{1,e}(c_1 \mid e)}{f_1(c_1 \mid e)}.$$

There is an additional term in the virtual cost that was not present before, which is very similar to the optimal sharing rule in Holmström (1979). This new term represents an additional reward for cost reduction that the developer receives through departures from bidding parity in the auction for production. This additional term serves to increase the sensitivity of the developer's virtual cost by increasing the marginal effect of a reduction in c_1 and thereby increasing $\phi_1(c_1)$. Furthermore, we know that the moral hazard term $F_{1,e}/f_1$ must be nonpositive, indicating that the developer is favored in the auction. Of course, the buyer realizes the developer did not shirk under the optimal scheme, but nevertheless the buyer must commit to "overreward" the developer for low costs if it wishes to maximize surplus from an *ex ante* point of view.

The additional term in the virtual cost of the developer reflects the interdependence of the moral hazard and adverse selection problems in this model. Rewards for low costs are accomplished by appropriately tilting the incentive scheme. Unlike Holmström, in our case rewards are made by changing the probability of winning the auction rather than through lump-sum payments since c_1 is not contractible.

5. CONCLUSIONS

The immediate implications of our analysis suggest that a policy of technology transfer is a useful device for reducing information rents in defense procurement. Indeed, it may be optimal to switch to a possibly inefficient bidder, *ex post*, to reduce rents, *ex ante*. Additionally, no information of the developer needs to be known by the second source for such a transfer to yield benefits for the government.

Although the contributions of this paper are not restricted to defense procurement, transferring information is not always possible in other auction contexts. For example, in the traditional private-values auction, the auctioneer cannot transfer the sub-

jective valuation of a painting from one bidder to another. Nonetheless, in many contexts such as defense procurement, the transfer of information is a real possibility because such information is embodied in tangible assets. As another example, following Laffont and Tirole (1988a), consider managerial takeover in this framework. Suppose that the incumbent managerial team secures profit for its stockholders following a particular profit plan. Later, a raider appears who may be employed to takeover the current management team and either institute its own profit strategies, or continue with its predecessors' plans (i.e., plans are transferable). In such a situation, takeovers may discipline incumbent management via threatened expropriation of managerial rents.[8]

Related to our work is that of Riordan and Sappington (1989). They consider a model of effort-enhanced value, in their no-commitment, unlimited-liability environment. Because the buyer cannot commit, the developer can expect the buyer to behave opportunistically after investment is sunk. Under this framework, the inability to commit not to use a second source leads to inefficient investment in most plausible cases. If commitment were possible, the government could promise to purchase the product at a price equal to its valuation and let the potential sellers bid away the expected information rents *ex ante* in the competition for the development contract at the symmetric information stage. In this paper, the limited-liability constraint implies that any gain from information rent reduction is a direct gain to the buyer. The tradeoffs involved are very different.

Laffont and Tirole (1988a) also consider a dynamic adverse-selection/moral-hazard framework. They find that if investment is completely transferable from the developer to the second source, the buyer would do best to commit to favor the developer at the competition for determining the producer. The results are similar in that bidding parity is disposed of to provide incentives for value-enhancing, transferable investment.

APPENDIX

Proof of Proposition 1: The proof of Proposition 1 proceeds with three lemmas. Lemma A.1 establishes necessary and sufficient conditions for truth-telling (Eq. (5)) and interim individual rationality (IIR), a weaker constraint than Eq. (6); the IIR constraint is given by:

$$\pi_j(c_j \mid c_j) \geq 0, \, \forall c_j, \, j-1,2.$$

---

[8] In the context of managerial incentives, Scharfstein (1988) examines the disciplinary role of a corporate raider who is informed of the firm's true value, and finds that such an informed raider both induces incumbent managers to work harder and reduces their information rents. His model is closely analogous to this paper in that the firm value (known by the incumbent managers) transfers completely to the raider if there is a takeover. This paper suggests that while a raider is more effective in reducing information rents if it knows the incumbent's information, there is nonetheless a positive role for uninformed raiders in reducing information rents. There is no requirement that the alternative agent have any *ex ante* knowledge of the primary agent's cost realization for information rents to be reduced.

Lemma A.2 establishes that the modified program of minimizing Eq. (4) over these new conditions is equivalent to solving P2 point-wise. Finally, Lemma A.3 shows that a particular solution to the modified program is "equivalent" to the solution of P1.

For notational convenience, we will sometimes denote a function that has had expectations taken over one argument, as a function of only the single remaining argument. For example, $\phi_1(c_1) \equiv \int_0^1 \phi_1(c_1, c_2) dF_2(c_2)$, etc.

Lemma A.1: Incentive compatibility (IC) and IIR hold if and only if

$$\pi_1(c_1 \mid c_2) = \pi_1(1 \mid 1) + \int_{c_1}^1 \phi_1(c_1) dc_1, \qquad (11)$$

$$\pi_2(c_2 \mid c_2) = \pi_2(1 \mid 1) + \int_{c_2}^1 [\phi_2(c_2) + \phi_0(c_2)\ell_2 dc_2, \qquad (12)$$

$$\phi_1(c_1) \geq \phi_1(c_1'), \ \forall c_1' > c_1, \qquad (13)$$

$$\phi_2(c_2) + \phi_0(c_2)\ell_2 \geq \phi_2(c_2') + \phi_0(c_2')\ell_2, \ \forall c_2' > c_2, \qquad (14)$$

$$\pi_i(1 \mid 1) \geq 0, \ i = 1, 2. \qquad (15)$$

Proof: Necessity: Consider firm 1. IC and the definition of $\pi_1(\hat{c}_1 \mid c_1)$ implies

$$\pi_1(c_1 \mid c_2) \geq \pi_1(\hat{c}_1 \mid c_1) = \pi_1(\hat{c}_1 \mid \hat{c}_1) - \phi_1(\hat{c}_1)(c_1 - \hat{c}_1).$$

Rearranging and reversing the roles of c_1 and $\hat{c}_1$ yields

$$-\phi_1(c_1)(c_1 - \hat{c}_1) \geq \pi_1(c_1 \mid c_1) - \pi_1(\hat{c}_1 \mid \hat{c}_1) \geq -\phi_1(\hat{c}_1)(c_1 - \hat{c}_1),$$

which implies Eq. (13). Without loss of generality, take $c_1 > \hat{c}_1$, divide by $(c_1 - \hat{c}_1)$, and take the limit as $c_1 \to \hat{c}_1$ to obtain

$$\frac{d\pi_1(c_1 \mid c_1)}{dc_1} = -\phi_1(c_1).$$

Since $\pi_1(c_1)$ is monotonic, it is Riemann integrable, thus implying Eq. (11). Finally, IR clearly implies Eq. (15). A similar series of arguments establishes the necessity of Eqs. (12), (14), and (15) for firm 2.

Sufficiency: Again consider firm 1. By definition of $\pi_1(\hat{c}_1 \mid c_1)$, we have

$$\pi_1(\hat{c}_1 \mid \hat{c}_1) = \pi_1(\hat{c}_1 \mid c_1) - \phi_1(\hat{c}_1)(\hat{c}_1 - c_1).$$

Condition (11) implies

$$\pi_1(c_1 \mid c_1) = \pi_1(\hat{c}_1 \mid c_1) + \int_{\hat{c}_1}^{c_1} [\phi_1(\hat{c}_1) - \phi_1(s)] \, ds.$$

But by condition (13), the integral is nonnegative, giving us incentive compatibility for firm 1. A similar series of arguments establishes the incentive compatibility for firm 2 using Eqs. (12), (14), and (15). IR follows immediately for both firms from conditions (11), (12), and (15).

Lemma A.2: The set of $\{\phi_i(c_1, c_2)\}_i$, which solves the modified IIR program, is the same as that which solves P2 below using point-wise minimization over $\{\phi_i\}_{i=0}^2$.

$$\min_{\phi_i} \left\{ \phi_1\left(c_1 + \frac{F_1(c_1)}{f_1(c_1)} \right) + \phi_2\left(c_2 + \frac{F_2(c_2)}{f_2(c_2)} \right) + \phi_0\left(\ell(c_1, c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right) \right\}.$$

Proof: The modified program is formally given by

$$\min_M \int_0^1 \int_0^1 \{t_i(c_1, c_2) + t_2(c_1, c_2)\} \, dF_1(c_1) \, dF_2(c_2)$$

subject to IC and IIR. Substituting out $t_i(c_1, c_2)$ in the objective function yields as the minimand

$$\int_0^1 \int_0^1 \{\pi_1(c_1, c_2) + \pi_2(c_1, c_2) + \phi_1(c_1, c_2)c_1$$
$$+ \phi_2(c_1, c_2)c_2 + \phi_0(c_1, c_2)\ell(c_1, c_2)\} \, dF_1(c_1) \, dF_2(c_2).$$

Integrating by parts and using Lemma A.1, we can simplify this expression to obtain the following objective function:

$$\min \int_0^1 \int_0^1 \left\{ \phi_1(c_1, c_2)\left(c_1 + \frac{F_1(c_1)}{f_1(c_1)} \right) + \phi_2(c_1, c_2)\left(c_2 + \frac{F_2(c_2)}{f_2(c_2)} \right) \right.$$
$$\left. + \phi_0(c_1, c_2)\left(\ell(c_1, c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right) + \pi_1(1 \mid 1) + \pi_2(1 \mid 1) \right\} \, dF_1(c_1) \, dF_2(c_2).$$

We want to minimize this subject to conditions (13) to (15). Rather than minimize subject to the monotonicity constraints, we will ignore them for now, and check our solution for their satisfaction.

Choosing the optimal $\phi_i(c_1, c_2)$ while ignoring the monotonicity constraints for the above integrand amounts to point-wise minimization of the bracketed expression over $\{\phi_i\}_i$.

To complete the lemma, we must show that the monotonicity conditions (13) and (14) hold. It is sufficient for monotonicity that $\phi_i(c_1, c_2)$ is nonincreasing in c_i and that both $\phi_2(c_1, c_2)$ and $\phi_0(c_1, c_2)$ are nonincreasing in c_2. Given $\ell_1 < 1$, and given our assumptions regarding the cost distributions, this is indeed the case.

Finally, we show that a solution to the relaxed IIR problem satisfies limited liability.

Lemma A.3: The following payments implement the optimal $\{\phi_i(c_1, c_2)\}_i$ for the relaxed IIR program and satisfy the limited liability constraints:

$$t_1(c_1, c_2) = \int_{c_1}^1 \phi_1(s)ds + \phi_1(c_1, c_2)c_1,$$

$$t_2(c_1, c_2) = \int_{c_2}^1 \{\phi_2(s) + \phi_0(s)\ell_2\}ds + \phi_2(c_1, c_2)c_2 + \phi_0(c_1, c_2)\ell(c_1, c_2).$$

Proof: Substituting the above payments into Eqs. (11) and (12) in the text demonstrates that the payments maintain incentive compatibility by Lemma A.1. Also, the payments clearly meet the limited liability constraint, as the integrals in the above expressions are never negative for any cost realization. Finally, there do not exist any other payments with lower expected value to the buyer. This last point is evident from Lemma A.2.

The transfers are determined directly from Lemma A.3.

Proof of Proposition 2: The result follows from noting that the gain from licensing is the expected reduction in virtual cost from licensing over a standard optimal auction without technology transfer. Since chosen virtual costs are only changed over Ω^0, we take expectations over this space. The expression immediately follows.

Proof of Lemma 1: The first-order condition for the solution is

$$\int_0^1\int_0^1 \pi_1(c_1, c_2) f_{1,e}(c_1 \mid e)dc_1 dF_2(c_2) - \psi'(e) = 0. \tag{16}$$

A sufficient condition for a maximum is that

$$\int_0^1\int_0^1 \pi_1(c_1, c_2) f_{1,ee}(c_1, e)dc_1 dF_2(c_2) - \psi''(e) < 0,$$

for all e. Integrating this expression by parts, and noting that Lemma A.1 from above implies $\partial\pi_1(c_1, c_2)/\partial c_2 = -\phi_1(c_1, c_2)$, yields an equivalent condition,

$$\int_0^1 [\pi_1(c_1, c_2)F_{1,ee}(c_1 \mid e)]_0^1 dF_2 + \int_0^1\int_0^1 \phi_1(c_1, c_2)F_{1,ee}(c_1 \mid e)dc_1 dF_2 - \psi''(e) < 0,$$

where e subscripts denote partial derivatives with respect to e. CDFC and the strict convexity of $\psi(e)$ assures us that the second-order condition for a maximum holds, thus Eq. (16) is both necessary and sufficient. Integrating by parts yields:

$$\int_0^1\int_0^1 -\frac{\partial\pi_1(c_1, c_2)}{\partial c_1}\frac{F_{1,e}(c_1 \mid e)}{f_1(c_1 \mid e)} dF_1(c_1 \mid e)dF_2(c_2) - \psi'(e) = 0.$$

Substituting from Lemma A.1 in the appendix gives us the desired result.

Proof of Proposition 3:   The moral hazard problem amounts to minimizing the expected cost of the buyer's expected payments, subject to the investment constraint given in Eq. (10). We can now summarize the new program as P3.

$$\min_{M',e} \int_0^1 \int_0^1 \left\{ \sum_{i=0}^1 \phi_i(c_1,c_2)\tilde{J}_i(c_1,c_2) - \psi(e) \right\} dF_1(c_1 \mid e) dF_2(c_2), \tag{17}$$

subject to monotonicity in $\phi_1(c_1)$ and $\phi_2(c_2)$ and to Eq. (10), where the $\tilde{J}_i(c_1,c_2)$ are the virtual types for the moral hazard problem as defined in the text. As before, we ignore the monotonicity constraints and check that our solution satisfies them.

Given our assumption that $\psi(1)=\infty$, we know by Eq. (10) that $e < 1$. Let μ be the Lagrange multiplier associated with the constraint in Eq. (10) and suppose for the moment that $\mu > 0$. Minimizing the Lagrangian taking the optimal choice of $\phi_1(c_1,c_2)$ as given, effort is chosen such that either

$$\mu \left\{ \int_0^1 \int_0^1 \phi_1(c_1,c_2) \left(\frac{F_{1,ee}(c_1 \mid \hat{e})}{f_1(c_1 \mid \hat{e})} \right) f_1(c_1 \mid \hat{e}) dc_1 dF_2(c_2) - \psi''(\hat{e}) \right\}$$

$$+ \int_0^1 \int_0^1 \phi_1(c_1,c_2) \frac{\partial}{\partial e} \left(\frac{F_1(c_1 \mid \hat{e})}{f_1(c_1 \mid \hat{e})} \right) f_1(c_1 \mid \hat{e}) dF_2(c_2)$$

$$= \psi'(\hat{e}) - \int_0^1 \int_0^1 \left[\sum_{i=0}^1 \phi_1(c_1,c_2)\tilde{J}_i(c_1,c_2) \right] f_{1,e}(c_1 \mid \hat{e}) dc_1 dF_2(c_2), \tag{18}$$

where $\tilde{J}_1(c_1,c_2) \equiv c_1 + F_1(c_1 \mid e)/f_1(c_1 \mid e) + \mu[F_{1,e}(c_1 \mid e)/f_1(c_1 \mid e)]$ and $\tilde{J}_0(c_1,c_2) \equiv$, $\ell(c_1,c_2) + \ell_2[F_2(c_2)/f_2(c_2)]$ or $e = 0$. By our assumptions on F and ψ, the marginal benefit from e is positive at $e = 0$, and so we know $e \in (0,1)$ and Eq. (18) holds.

Now, given that e is optimally set at $\hat{e}$ and given the value of $\mu > 0$, we may solve for the optimal $\phi_i(c_1,c_2)$. Bringing the investment constraint within the objective function yields

$$\int_0^1 \int_0^1 \left\{ \sum_{i=0}^1 \phi_i(c_1,c_2)\tilde{J}_i(c_1,c_2) \right\} df_1(c_1 \mid \hat{e}) dF_2(c_2) - \mu\psi'(\hat{e}) - \psi(\hat{e}).$$

But the solution to the minimum of this expression is identical as the point-wise minimum of

$$\min_{\phi_i} \left\{ \phi_1 \left[c_1 + \frac{F_1(c_1 \mid \hat{e})}{f_1(c_1 \mid \hat{e})} + \mu \frac{F_{1,e}(v \mid \hat{e})}{f_1(c_1 \mid \hat{e})} \right] + \phi_0 \left[\ell(c_1,c_2) + \ell_2 \frac{F_2(c_2)}{f_2(c_2)} \right] \right\}.$$

The problem is therefore as in the proposition. Providing that $\mu > 0$, the virtual costs are appropriately monotone in costs so as to satisfy the additional monotonicity constraints. Finally, $\mu > 0$ holds, since the developer will ignore the positive externality that effort has on reducing licensed costs. Thus, the purchaser will always prefer

more effort than the developer is willing to provide, and the constraint must bind with a positive multiplier.

CONSTRUCTING INCENTIVE SCHEMES FOR GOVERNMENT CONTRACTS: AN APPLICATION OF AGENCY THEORY
by Stefan Reichelstein[1]

1. INTRODUCTION

In recent years, the agency paradigm has become central to theoretical research in managerial accounting. Empirical research in the area of executive compensation has tested the consistency of agency model predictions with observed compensation data.[2] However, it seems difficult to trace specific instances where results and insights obtained from agency models have affected actual management practice.[3] This paper describes one such instance in the context of government contracting, showing how agency theory was used to design incentive contracts.

The research reported here was initiated by the German Department of Defense (GDOD). The department commissioned a study to examine the applicability of a class of incentive schemes subsequently referred to as *budget-based schemes*.[4] To implement these schemes, the GDOD expressed interest in a constructive procedure that would derive suitable budget-based schemes for specific procurement projects. I describe such a procedure and discuss a number of institutional factors that affected the way the budget-based schemes were applied to two pilot projects in Germany.

[2]See, e.g., Lambert and Larcker (1987), Antle and Smith (1986), and Rosen (1990). Wolfson (1985) examines the extent to which sharing rules employed by oil and gas partnerships reflect agency problems.

[3]Johnson and Kaplan (1987, p. 175) express particular pessimism about the practical relevance of agency theory. They state that, "because of mathematical difficulties and the consequent lack of realism of the models and their assumptions, the potential for agency theory research to inform or even understand managerial accounting practice will likely be realized, if at all, only many years in the future."

[4]The budget-based schemes belong to the larger category of menus of linear contracts, studied extensively in the literature (see, e.g., Laffont and Tirole, 1986; McAfee and McMillan, 1987; Rogerson, 1987; Melumad and Reichelstein, 1989; Picard, 1987; and Baron and Besanko, 1987b, 1988). The models in those studies pertain to a variety of economic contexts, including procurement contracting, regulation, and performance evaluation of managers within firms.

Traditionally, government contracts under sole-source conditions have been awarded either as fixed-price or cost-plus contracts. The use of fixed-price contracts has been confined to projects with relatively few technological and economic uncertainties. With such uncertainties, fixed-price contracts are unattractive from a risk-sharing perspective. Yet, even with a risk-neutral contractor (because of size and diversification) governments are typically reluctant to sign fixed-price contracts when there are major informational asymmetries. If the government is relatively ignorant about inputs and resources required for the project, it will have difficulties disputing the firm's *ex ante* cost calculation. As a consequence, the firm earns an informational rent; that is, the firm will extract a higher price than it would have if the government had shared the firm's knowledge and expertise.

Cost-plus contracts avoid the problem of overpayment, but, as has been well documented, the government subjects itself to the problem of cost padding. To limit the negative incentives of cost-plus contracts, it has become common practice in the United States to replace standard cost-plus contracts with cost-plus-fixed-fee contracts, so that the firm's profit allowance is fixed rather than being proportional to actual project costs.[5] In recent years, there has been an increasing trend in the United States to provide positive incentives for cost control by using cost-plus-incentive-fee contracts.[6] At the outset of a project, the parties negotiate a cost target, and the firm's profit increases proportionally with cost underruns relative to the cost target. Conversely, the incentive profit decreases at the same rate with cost overruns. In effect, the firm thus bears a share of actual project costs. A recent survey by the U.S. General Accounting Office (GAO, 1987) shows that for the period 1978–1984 firms' cost-share parameters typically varied between 15 and 25 percent but were as high as 50 percent in unusual cases.

A major concern voiced repeatedly in connection with cost-plus-incentive-fee contracts is that the government is unable to formulate realistic cost targets for many projects. If the target is set unrealistically low, the firm is likely to suffer a financial penalty. Conversely, an unrealistically high cost target leads to additional "undeserved" profits. For this reason, the GAO states that cost-plus-incentive-fee contracts are confined to procurement projects where "the government has a sound basis to estimate contract costs, but where uncertainties exist that make a fixed-price contract impractical" (GAO, 1987, p. 1). The relatively low cost-share parameters currently used (15 to 25 percent) may reflect the government's desire to mitigate the effects of unrealistic cost targets.

The budget-based schemes considered in this study can be viewed as a refinement of cost-plus-incentive-fee contracts. In addition to actual cost, the incentive fee now depends as well on a cost estimate that the firm submits, typically at the start of the

---

[5]However, according to former Under Secretary of Defense R. Fox, generous provisions for reimbursement of "unusual" expenses create conditions that closely resemble standard cost-plus arrangements (see Fox, 1974). In Germany, cost-plus-proportional-fee contracts are currently used, though with increasing controversy.

[6]McAfee and McMillan (1988) summarize the evolution of incentive schemes used by the U.S. Navy for shipbuilding contracts in the 1960s and 1970s. In Germany, cost-plus-incentive-fee contracts have been used only for exceptional projects until recently.

project. In effect, the firm selects a budget (target cost), and the incentive profit is proportional to the budget variance. Previous modeling analysis has shown that budget-based schemes create desirable reporting and performance incentives (see Kirby et al., 1991). The government receives information that is useful for its budget planning process, since the contracting firm is induced to submit an unbiased cost estimate. Specifically, the firm has an incentive to reveal truthfully its own assessment of expected project costs. To some extent, the budget-based schemes therefore avoid the issue faced by cost-plus-incentive-fee contracts described above. Instead of having the government formulate a realistic cost target, this task is now left to the better informed firm.

From a cost-control perspective, the budget-based schemes have been shown to be optimal incentive mechanisms. By offering a menu of contracts, the government can tailor performance incentives to the firm's privately observed cost information.[7] Specifically, the firm chooses a high target profit in return for a high cost-share parameter, provided its cost information is relatively favorable. A high cost-share parameter will induce the firm to conduct the project in a more efficient way. The resulting cost savings are effectively split, since the firm receives a large incentive profit. As a consequence, both sides will be better off.

The following section of this paper contains a summary of the relevant incentive literature. In particular, I review the reporting and performance incentives created by the budget-based incentive schemes. A procedure that enables the government to construct a suitable budget-based scheme for a given procurement project is described in Section 3. This procedure requires the government to assess a number of parameters reflecting the underlying agency problem and then carry out a sequence of computations to obtain the desired incentive contract.

Section 4 reports on applications of the budget-based schemes in Germany. The discussion first focuses on a number of modifications and constraints that representatives of the GDOD formulated to make the schemes compatible with current procurement policy. One budget-based scheme that is currently being used in a pilot project is described along with the criteria that led to its selection. Section 5 concludes with a brief discussion of the potential use of budget-based schemes for regulating public utilities.

2. THEORY BACKGROUND

Suppose that a government wants to conduct a project for which there is only one viable supplier (e.g., the government faces a sole-source contract). Frequently, both parties will be uncertain about the necessary cost for the project, yet the contracting firm has better cost information. In such cases the government will typically use a cost-based contract. The project price then becomes the sum of actual costs incurred, as verified by the government's auditors, plus a profit allowance for the con-

---

[7]The advantage of a menu of contracts over a single incentive contract has been analyzed in the managerial accounting literature under the heading "value of communication" (see Baiman and Evans, 1983, Melumad and Reichelstein, 1989, and Penno, 1984).

tracting firm. Under a cost-plus-fixed-fee contract, the firm's profit is a constant, yet it is a decreasing function of actual cost under a cost-plus-incentive-fee contract. With actual cost denoted by x, the price paid by the government can be expressed as:

$$P = x + H(x),$$

with

$$H(x) = a + b \bullet (T - x). \tag{1}$$

The contractor's profit, $H(x)$, consists of a target profit a, plus a bonus (penalty) for cost underruns (overruns) relative to a prespecified cost target T. The parameter b $(0 \leq b \leq 1)$ is usually referred to as the cost-share parameter. A cost-plus-fixed-fee contract results if $b = 0$.

In the present study, attention is restricted to incentive-fee functions of the form:

$$H(E,x) = a(E) + b(E) \bullet (E - x). \tag{2}$$

The contracting firm now determines the target cost through its estimate E. Deviations from this standard are rewarded or penalized at the rate $b(E)$, which itself depends on the firm's cost estimate. Similarly, it will be necessary to vary the target profit, $a(E)$, with the cost estimate to create appropriate reporting incentives. Effectively, the government offers a menu of contracts. By submitting its cost estimate, the firm chooses one particular incentive fee function (linear in x) from the menu. Subsequently, the schemes in Eq. (2) will be referred to as *budget-based schemes*.

From the government's perspective, a menu of contracts may be preferable to a single incentive contract for two reasons. First, a menu may conceivably entail better performance incentives, that is, more effective incentives for cost control. Second, a menu of contracts allows the government to elicit cost information from the better informed firm. Ignoring the issue of cost control, it can be shown that the budget-based schemes induce the firm to provide an unbiased estimate of expected project cost, provided the functions $a(\bullet)$ and $b(\bullet)$ in Eq. (2) are suitably calibrated. Specifically, consider the following restrictions:[8]

$$a(E) \text{ is a convex and decreasing function and} \tag{3}$$

$$b(E) = -a'(E). \tag{4}$$

Suppose the firm subjectively assesses the expected project cost to be z.[9] If it reports E under a budget-based scheme, the expected profit becomes $a(E) + b(E) \bullet (E - z)$. In

---

[8]If $a(\bullet)$ is not differentiable at E, then $a'(E)$ should be interpreted as a subgradient of $a(\bullet)$ at E.

[9]Formally, the firm's subjective beliefs regarding the uncertain cost x are represented by a probability distribution $F(\bullet)$ whose mean is z, i.e.,

$$z = \int x dF(x).$$

particular, the expected fee equals $a(z)$ if the true expectation z is reported. Incentive compatibility requires that:

$$a(z) \geq a(E) + b(E) \bullet (E - z), \tag{5}$$

for arbitrary values of z and E. After substituting Eq. (4) into Eq. (5), the self-selection condition becomes:

$$a(z) \geq a(E) - a'(E) \bullet (E - z) \geq 0, \tag{6}$$

which will be satisfied provided that the function $a(\bullet)$ is convex. It can be shown furthermore (see Osband and Reichelstein, 1985) that the budget-based schemes are essentially unique in eliciting the mean of the privately known distribution $F(\bullet)$. To induce the same cost report for *all* distributions $F(\bullet)$ with mean z, the profit function must be linear in the actual cost variable x.[10]

Consider now the issue of performance incentives. There is a common perception that cost-based government contracts are subject to cost-control problems. Firms may be tempted to pad project costs through expanded purchases of equipment, excessive testing and experimentation, or generous arrangements with their suppliers.[11]

It has been argued that the method for determining project costs can itself create undesired incentives. A major issue here is the application of overhead costs. Consider a situation where a government project and other commercial products share common production facilities. If overhead costs are allocated to the government project on the basis of direct labor hours, the firm may have an incentive to increase the number of direct labor hours for the government project, because direct labor costs are reimbursed and every hour of additional labor shifts overhead costs from the commercial products to the government contract.[12] Without countervailing incentive provisions, the firm will increase its profit by expanding the number of direct labor hours for the government project.[13] Rogerson (1992d) develops a model that

---

[10]Note that the budget-based schemes differ from the schemes examined by Weitzman (1976) in connection with a Soviet incentive system (Kaplan and Atkinson, 1989, refer to them as "truth-inducing" schemes). Because the schemes discussed there consist of a menu of piece-wise linear contracts, the informed agent (contracting firm) can be induced to report a given quantile of the distribution $F(\bullet)$ (e.g., the median). However, in general, such schemes cannot elicit the mean.

[11]Representatives of the GDOD mentioned the following specific cost-padding problem. Suppose a contracting firm finds that it has hired too many workers. The firm can reduce its workforce but then it will incur certain layoff costs, such as severance pay and social security payments. Alternatively, the firm can assign more workers than are actually needed to a government contract. Effectively, the firm pads the number of labor hours for the project and is better off as long as project costs are fully reimbursed.

[12]This argument presumes that actual overhead costs do not increase "too fast" as direct labor hours increase.

[13]The same argument applies if the government adopts a normal costing format, i.e., direct costs are reimbursed at actual levels, while overhead costs are applied in proportion to some allocation base such as direct labor hours. If *actual* overhead costs increase at a rate less than the overhead application rate, the firm will be able to cover a larger share of its actual overhead cost by expanding the direct cost factors beyond the necessary level.

quantifies the distortions associated with overhead allocation rules commonly employed in defense contracting.

To model the possibility of cost padding, suppose the firm finds initially that the minimal expected project cost is μ. If the firm engages in the type of cost-increasing activities described above, it chooses a "slack" parameter $\alpha \geq 0$, with $\alpha = 0$ representing maximum efficiency. Since α measures the dollar amount of slack, it is natural to suppose that the expected project cost increases to $z = \mu + \alpha$. Actual cost, x, remains a random variable whose mean becomes $z = \mu + \alpha$. The firm's pecuniary benefit from incorporating slack is denoted by $B(\alpha)$, which represents the dollar value associated with slack level α.

In designing an incentive contract, the government faces the following tradeoff. Because the firm has better cost information (the parameter μ), stronger incentives for cost control can be obtained only at the expense of higher profit. As mentioned above, fixed-price contracts would induce perfect incentives to control project costs, yet the government is likely to overpay because the fixed price would be calculated so that the firm earns a minimum acceptable level of profit, even in case of high project costs. To balance the tradeoff between stronger efficiency incentives and higher firm profits, it will be optimal in general to adopt an incentive scheme that results in a limited amount of slack.

Given any incentive contract, let $\alpha(\mu)$ denote the choice of slack that maximizes the firm's expected profit when its cost environment is μ. From the government's perspective, the optimal slack policy $\alpha(\mu)$ is determined by the benefit function $B(\bullet)$ as well as by prior beliefs regarding the firm's cost environment μ. Those beliefs are represented by a probability distribution function $N(\mu)$ with support $[\underline{\mu}, \overline{\mu}]$ and density function $n(\mu)$. A common assumption in models like the present one is that the "inverse hazard rate":

$$h(\mu) \equiv \frac{N(\mu)}{n(\mu)}$$

be monotone increasing.[14] The following result is taken from Kirby et al. (1991).

Lemma 1: If $h(\mu)$ is increasing in μ and the benefit function $B(\bullet)$ is concave with nonnegative third derivative, then a budget-based scheme of the form shown in Eqs. (2) through (4) is an optimal incentive contract.

The conditions of the lemma imply that an optimal incentive scheme induces a separation of types. Unlike low-cost types (low μ), high-cost types pad project costs to a larger extent (i.e., the function $\alpha[\mu]$ is increasing in μ).[15] The budget-based schemes in Eqs. (2) through (4) induce exactly that separation, since higher cost estimates imply lower cost-share parameters, which in turn means a lesser incentive to reduce

---

[14]This condition is satisfied by many of the commonly considered probability distributions (see, e.g., Laffont and Tirole, 1991, and the references contained therein).

[15]This result has been established by Laffont and Tirole (1986), McAfee and McMillan (1987), and Rogerson (1987).

slack. Given a budget-based scheme, the firm's expected payoff (i.e., profit plus benefit derived from slack), becomes:

$$a(E) + b(E) \cdot [E - (\mu + \alpha)] + B(\alpha). \tag{7}$$

The contracting firm seeks to maximize this expression with respect to α and E. Provided the budget-based scheme satisfies conditions (3) and (4) above, the firm will always report $E = \mu + \alpha$, regardless of the value of α to be chosen. Hence, whatever amount of slack, α, the firm chooses, it will provide an unbiased estimate of expected project costs. As a consequence, maximization of Eq. (7) reduces to maximizing $a(\mu + \alpha) + B(\alpha)$, with regard to α. For the contracting firm to choose slack $\alpha(\mu)$ when the cost environment is μ, it is necessary that:

$$a'[\mu + \alpha(\mu)] + B'[\alpha(\mu)] = 0, \tag{8}$$

provided $\alpha(\mu) > 0$. Because of Eq. (4), Eq. (8) can be rewritten as:

$$b[\mu + \alpha(\mu)] = B'[\alpha(\mu)]. \tag{9}$$

Equation (9) will be used to determine the cost-share parameters in the next section. However, for this equation to become operational, the government first needs to identify the desired slack policy $\alpha(\mu)$.

Lemma 1 confirms the intuition that it is preferable for the government to let the better-informed firm determine the cost-share parameter. With a traditional cost-plus-incentive-fee contract, the firm will choose the same amount of slack in all cost environments so that the marginal value of slack equals the cost-share parameter of the incentive contract. Under a budget-based scheme, the government can additionally offer incentive contracts with higher cost-share parameters. The firm will select a "steeper" contract if it has favorable cost information. In those cases, the firm will earn a higher profit, but since the project is conducted more efficiently, both sides can be made better off.

3. CONSTRUCTING BUDGET-BASED SCHEMES

From a practitioner's perspective, implementation of a cost-plus-incentive-fee contract requires the choice of three numbers: a cost-share parameter, a target cost level, and a target profit level.[16] In contrast, the budget-based schemes, which effectively consist of a menu of cost-plus-incentive-fee contracts, require an entire schedule of these parameters. This leaves the government with the task of selecting budget-based schemes for alternative procurement projects. Preferably, the selection procedure should not require detailed understanding of the underlying incentive theory.

---

[16]GAO (1987) describes how the government determines these parameters.

To state a specific budget-based contract, the government may use the general formula in Eq. (2) in conjunction with formulas for the functions $a(\bullet)$ and $b(\bullet)$. A simpler alternative may be to express the contract by tabulating the incentive profit that the firm will receive for alternative levels of estimated and actual costs.[17]

The following steps provide a procedure for constructing a suitable budget-based incentive scheme. The government first assesses a number of basic parameters that reflect its beliefs regarding minimum cost and cost-padding opportunities. On the basis of these parameters, the government carries out a sequence of computations whose final output is the desired matrix of incentive profits. This matrix can be supplemented with a set of interpolation rules that compute the incentive profit for values of estimated and actual cost not contained in the matrix.

Step 1: Assessing Basic Parameters

First, the government has to assess the interval $[\underline{\mu}, \overline{\mu}]$; that is, the range of possible cost expectations that the firm might hold if it conducted the project with maximum efficiency. For most projects, the government will formulate a reservation or ceiling value representing the highest cost estimate (unbiased) for which it would be willing to undertake the project. When this ceiling value is denoted by $\overline{E}$, the range of admissible cost estimates becomes the interval $[\underline{\mu}, \overline{E}]$. In the subsequent analysis, $\overline{E}$ is considered an exogenously given parameter.[18]

To construct the matrix of incentive profits, it will be useful to divide the interval $[\underline{\mu}, \overline{E}]$ into k subintervals. The interpretation is that the firm can choose one of the $(k + 1)$ estimates E_i, where $E_i < E_{i+1}$ for $1 \le i \le k$. It will be convenient to let $\underline{\mu} = E_0$ and $\overline{E} = E_k$. An obvious tradeoff is associated with the number k. The larger the value of k, the closer the discrete incentive scheme will approximate the optimal continuum solution.[19] On the other hand, too large a value of k would make the resulting matrix of incentive profits impractical. Representatives of the GDOD felt that k should be between five and ten (in the application described in Section 4, k is equal to 7).

Step 2: Determining the Cost-Share Parameter $b(\overline{E})$

As argued in the previous section, the optimal cost-share parameters are given, in principle, by Eq. (9) with the optimal slack policy $\alpha(\mu)$ determined by the two characteristics of the underlying agency problem: the government's prior distribution

---

[17]This approach has been implemented by IBM in the context of sales force compensation. According to Gonik (1978), IBM offered a menu of compensation plans via a matrix, whose entries show a manager's sales commission for alternative combinations of forecasted and actual sales. Also, Gates (1988) describes a contracting situation where an "agent" was offered a menu consisting of two alternative incentive schemes.

[18]The ceiling $\overline{E}$ can be derived as follows. Suppose the government attaches a gross benefit of R dollars to the project. For R sufficiently large, the project will be conducted for any cost environment μ. Furthermore, the lemma in Section 2 states that a budget-based scheme is optimal. It can be shown that, as R decreases, a "truncated" budget-based scheme is optimal, such that the government conducts the project only if $E \le \overline{E}$, where $\overline{E}$ is a decreasing function of R.

[19]In the context of a monopolist offering a menu of services, Wilson (1988) examines the loss associated with offering just k different services rather than a continuum. Wilson shows that the relative loss associated with a finite menu goes rapidly to zero as k gets large. Specifically, the difference between the continuum and the finite solution converges to zero at a rate of at least $1/k^2$.

$N(\mu)$ and the benefit function $B(\alpha)$.[20]   To characterize the optimal function $b(\bullet)$, specific functional forms for $N(\mu)$ and $B(\alpha)$ are considered. The resulting characterizations are useful only to the extent that the functional forms provide a "reasonable" approximation of the actual functions $N(\mu)$ and $B(\alpha)$ assessed by the government.

In considering the range of desired cost-share parameters, it is possible to provide upper and lower bounds for the cost-share parameters in terms of the firm's benefit function, $B(\alpha)$.   Recall from Section 2 that $B(\bullet)$ is assumed to be concave with $0 \leq B'(\alpha) \leq 1$.   Denote $\upsilon_1 = B'(0)$. A special case of interest occurs if the function $B(\bullet)$ is such that from a certain point on, the marginal value of slack remains constant (e.g., the firm receives the equivalent of 20 cents for every additional dollar of slack).   Formally:

$$B'(\alpha) = \upsilon_2 \ \text{ for } \alpha \geq \alpha^*. \tag{10}$$

Observation 1: If Eq. (10) holds, the optimal cost-share parameters satisfy:

$$\upsilon_2 \leq b(E) \leq \upsilon_1. \tag{11}$$

As shown in the appendix, the upper bound υ_1 is indeed the optimal sharing parameter corresponding to the lowest estimate. As a consequence, the firm will conduct the project with maximum efficiency (i.e., $\alpha = 0$). Note that if $\upsilon_1 = 1$, the incentive contract assigns a fixed-price contract to the lowest cost estimate. The result $b(E) \geq \upsilon_2$ reflects that the government wants to contain the amount of slack at α^*.[21]

As argued above, incentive compatibility requires the function $b(\bullet)$ to be monotone decreasing. Characterizing the curvature of $b(\bullet)$ turns out to be difficult in general, but a tractable special case results when the benefit function $B(\bullet)$ is of the following linear-quadratic form:

$$B(\alpha) = \begin{cases} s_1 \bullet \alpha - \dfrac{s_2}{2} \bullet \alpha^2 & \text{if } \alpha \leq \alpha^* \text{ and} \\[2mm] (s_1 - s_2 \bullet \alpha^*) \bullet \alpha + \dfrac{s_2}{2} \bullet \alpha^{*2} & \text{if } \alpha > \alpha^*. \end{cases} \tag{12}$$

The three parameters (s_1, s_2, α^*) are to be selected so that $s_1 - s_2 \bullet \alpha^*$ is positive, which ensures that $B(\bullet)$ is increasing and concave. Note that, by definition, $s_1 - s_2 \bullet \alpha^* = \upsilon_2$. As shown in the appendix, the desired slack policy $\alpha(\mu)$ will be increasing at the same rate as the inverse hazard rate, $h(\mu)$, if the benefit function is of the linear-quadratic form in Eq. (12). Hence $\alpha(\mu)$ will be convex (concave) whenever $h(\mu)$ is convex (concave).

To tie this characterization to the curvature of $b(\bullet)$, note that the right-hand side of Eq. (9) is linearly decreasing in α for linear-quadratic benefit functions. Thus, one

---

[20]The exact formula for the optimal $\alpha(\mu)$ in terms of $N(\mu)$ and $B(\alpha)$ is given by Eq. (A.1) in the appendix.

[21]An implicit assumption of the underlying agency model is that the presence of the government's auditors does not create "effective" upper bounds on the amount of slack that the firm could choose.

would expect $b(\bullet)$ to be decreasing at a decreasing rate (i.e., $b(\bullet)$ to be convex) whenever $\alpha(\mu)$ increases at a decreasing rate. To state the formal result, it is useful to introduce the following terminology: The relevant domain of the function $b(\bullet)$ is the interval (E_*, E^*), where E_* is defined to be the largest value of E such that $b(E) = \upsilon_1$ for all $E \leq E_*$, and E^* is the smallest value of E such that $b(E) = \upsilon_2$ for $E \geq E^*$. Hence the relevant domain is the range of cost estimates where the function $b(\bullet)$ assumes values intermediate between υ_1 and υ_2 (of course, it is possible that $E_* = \underline{\mu}$, and $E^* = \overline{E}$).

Observation 2: If Eqs. (10) and (12) hold, the function of optimal cost-share parameters is convex (concave) on the relevant domain, whenever the function $h(\mu)$ is concave (convex).[22]

An immediate consequence of this result is that when the government's prior distribution is uniform, the cost-share parameters decrease linearly on the relevant domain. The appendix contains a proof of Observation 2 and an explicit computation of the optimal sharing parameters for the following family of probability densities:

$$n(\mu) = \begin{cases} c \bullet (\mu - \underline{\mu})^r & \text{if } \mu \leq m \text{ and} \\ c \bullet (\overline{\mu} - \mu)^r & \text{if } \mu \geq m. \end{cases} \tag{13}$$

Here, $r \geq 0$ is a parameter to be chosen, m is the midpoint between $\underline{\mu}$ and $\overline{\mu}$, i.e.,

$$m = \frac{1}{2}(\underline{\mu} + \overline{\mu}),$$

and c is a constant ensuring that $n(\bullet)$ is a density.[23] If $r = 0$, the distribution is uniform; for $r > 0$, the distribution is single peaked and symmetric around the mean m. The parameter value r indicates the variance of the random variable μ. The higher the value of r, the lower the variance, since more probability weight is centered around the mean m. Figure 7.1 illustrates the shape of this distribution for four alternative values of r ranging between zero and 1.5.

For the family of probability distributions in Eq. (13), it turns out that the inverse hazard rate, $h(\mu)$, is linear on the interval $[\underline{\mu}, m]$ and strictly convex thereafter. In light of Observation 2, the function $b(\bullet)$ is therefore concave on the relevant domain. The explicit formula for $\underline{b}(\bullet)$ in terms of the six parameter values ($\underline{E}, \overline{E}, s_1, s_2, \alpha^*, r$) is given in the appendix.

---

[22]It is sufficient that $h(\mu)$ be concave (convex) for those μ corresponding to the relevant domain, i.e., on the interval [E_*, u^*], where $\mu^* + \alpha^* = E^*$

[23]It is readily verified that $N(\overline{\mu}) = 1$ if,

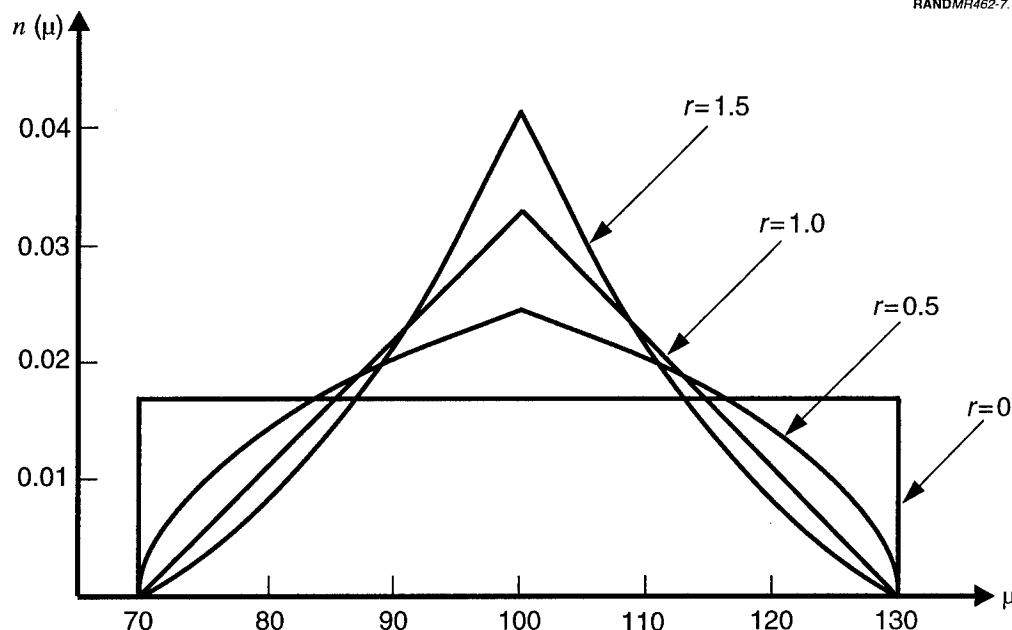$$c = \frac{r+1}{2}(m - \underline{\mu})^{-(r+1)}.$$

Figure 7.1—Illustration of the Density Given by Eq. (13)

In summary, the above analysis provides a partial characterization of the desired function $b(\bullet)$ with regard to its range and curvature. An explicit computation of the optimal cost-share parameters is possible for the specific functional forms given in Eqs. (12) and (13). It should be noted, though, that these functional forms may not provide a "sufficiently good" approximation of the functions $N(\bullet)$ and $B(\bullet)$ as assessed by the government for a particular procurement project. For example, the government's prior distribution may not be symmetric around the mean, and therefore the functional form in Eq. (13) may not provide an appropriate representation. In those cases, a contract designer may seek to access the underlying agency model directly. For given functions $N(\bullet)$ and $B(\bullet)$, the optimal slack policy $\alpha(\mu)$ can be derived according to Eq. (A.1) in the appendix. Thereafter, the optimal cost-share parameters $b(E)$ can, in principle, be determined according to Eq. (9). Most likely, these solutions would be obtained in practice through numerical analysis.

To derive the matrix of incentive profits, it is, of course, sufficient to select only the $k + 1$ cost-share parameters $b(E_i)$, $0 \leq i \leq k$. It will be convenient in the subsequent steps to proceed as if the desired function $b(\bullet)$ is linear on each interval (E_i, E_{i+1}). One then obtains a piece-wise linear approximation of the "optimal" $b(\bullet)$. The loss associated with this approximation diminishes as k gets larger.

Step 3: Computing the Target Profits $a(E)$

The discussion in Section 2 shows that, up to a constant of integration, the target profits $a(E)$ of any incentive compatible budget-based scheme are uniquely determined by the cost-share parameters $b(E)$. Hence, the only discretion left to the con-

tract designer is a boundary value, denoted by c, that gives the target profit for the highest cost estimate $E_k = \overline{E}$. Thus,

$$a(\overline{E}) = c. \tag{14}$$

If the threshold value is not to distort the reporting incentives, the constant c must satisfy the requirement that the firm is indifferent between accepting and rejecting the contract when its estimate of expected project cost equals $\overline{E}$. As will become clear in the next section, however, the choice of c is likely to be affected by other constraints in practice.

Observation 3: Given a function of cost-share parameters $b(\bullet)$, which is linear on each interval (E_i, E_{i+1}), the target profits can be computed recursively by Eq. (14) and:

$$a(E_i) = a(E_{i+1}) + \frac{1}{2}[b(E_i) + b(E_{i+1})] \bullet (E_{i+1} - E_i), \quad \text{for } 1 \leq i \leq k-1. \tag{15}$$

To verify Eq. (15), recall that Eq. (4) implies:

$$a(E_i) = a(E_{i+1}) + \int_{E_i}^{E_{i+1}} b(E)dE, \tag{16}$$

for arbitrary values of E_i and E_{i+1}. Since the cost-share parameter function $b(\bullet)$ is linear on each segment $(E_i \, E_{i+1})$, an immediate geometric argument shows that the integral of $b(\bullet)$ between E_i and E_{i+1} is given by

$$\frac{1}{2}\Big[b(E_{i+1}) + b(E_i)\Big] \bullet (E_{i+1} - E_i).$$

Hence, Eq. (15) holds.

To complete the construction of the matrix of incentive profits, one needs to choose p different actual cost figures $\{x_1, \ldots, x_p\}$. Since the firm has an incentive to provide an unbiased report of the expected cost in the first stage, it seems natural to choose the values x_1 from the range of possible cost estimates, that is, the interval $[\underline{\mu}, \overline{E}]$. Given values for $b(E_i)$ and $a(E_i)$, the matrix of incentive profits can thereafter be calculated according to Eq. (2) for all $k \bullet p$ pairs of estimated and actual cost.

Step 4: Interpolation Rules

Representatives of the GDOD suggested that the matrix of incentive profits be appended by interpolation rules that determine the incentive profit for combinations of x and E not contained in the matrix. With regard to actual cost x, the interpolation rules are obvious, since the budget-based schemes are linear in actual cost. Therefore, if $x = \beta \bullet x_j + (1-\beta) \bullet x_{j+1}$, the incentive profit would be interpolated linearly:

$$H(E_i, x) = \beta \cdot H(E_i, x_j) + (1 - \beta) \cdot H(E_i, x_{j+1}). \tag{17}$$

Similarly, the incentive profit would be extrapolated linearly for actual cost levels x outside the interval $[x_1, x_p]$.

With regard to the estimates E, the interpolation rules have to create a new row in the matrix for any estimate E with $E_i < E < E_{i+1}$. This requires calculation of $a(E)$ and $b(E)$ in terms of the neighboring values, E_i and E_{i-1}. Since the function $b(\cdot)$ is taken to be linear on the interval $[E_i, E_{i+1}]$, the natural interpolation rule is:

$$b(E) = b(E_{i+1}) + \frac{b(E_i) - b(E_{i+1})}{E_{i+1} - E_i} (E_{i+1}) - E), \tag{18}$$

and

$$a(E) = a(E_{i+1}) + \frac{1}{2} [b(E) + b(E_{i+1})] \cdot (E_{i+1} - E). \tag{19}$$

These equations can be combined with the basic formula in Eq. (2) to generate the new row corresponding to any estimate E with $E_i < E < E_{i+1}$. Note that if the functional $b(\cdot)$ is linear on each interval (E_i, E_{i+1}), then the expressions in Eqs. (18) and (19) generate the values that would have resulted had the estimate E been one of the points of the partition. Thus, if the matrix of incentive profits is supplemented by the interpolation rules in Eqs. (17) through (19), it will lead to the same behavior and performance as a continuum scheme whose function $b(\cdot)$ is piece-wise linear.

4. APPLICATION OF THE BUDGET-BASED SCHEMES IN GERMANY

Profit Policy

When the budget-based schemes were presented to representatives of the GDOD, the first concern was to ensure that these schemes were compatible with existing procurement rules. The German procurement laws set a framework for the calculation of contractors' profits under cost-based contracts.[24] The contractor's base profit is typically proportional to the value of the project as measured by actual costs verified *ex post* by government auditors.[25] The markup factor currently amounts to about 5 percent, though this varies slightly by the extent to which project components are supplied by subcontractors.[26]

---

[24] The general rules are stated in the so-called "Leitsätze zur Preisverordnung," abbreviated as LSP.

[25] As mentioned in the introduction, such cost-plus-proportional-fee contracts are not admissible in the United States. Currently, there are various efforts under way to change the formula for the calculation of base profits in Germany.

[26] As discussed by Osband (1988b) and Rogerson (1992a) the calculation of base profit in the United States is concerned with two major considerations: that the contracting firm earn an "adequate" return on capital and, in addition, be reimbursed for certain unrecognized costs, e.g., interest expenses.

In addition to base profit, contractors may be paid an incentive or performance profit. Although there is no general formula for the calculation of such profits, the law stipulates that incentive profits may be granted only to reward "exceptional contractor effort with regard to technical, organizational or economic performance."[27] The question then becomes whether the incentive profits arising from the use of a budget-based scheme in fact reward exceptional contractor effort. For instance, if the firm submits a relatively low estimate E and actually achieves that standard (i.e., $x = E$), it will be entitled to an incentive profit of $a(E)$, even though, relative to the self-selected standard E, there was no "exceptional contractor effort." It was agreed that the budget-based schemes are nonetheless compatible with the German procurement laws. By submitting a low estimate, the firm subjects itself to a larger risk of negative incentive profits and this exposure warrants a premium, even if the firm merely achieves its self-selected standard.[28] To embed the budget-based incentive schemes into the current procurement framework, it was suggested that the incentive profit be viewed as a payment over and above the firm's base profit. Accordingly, the price paid by the government can be expressed as:

$$\text{Price} = \text{Actual Cost} + \text{Base Profit} + \text{Incentive Profit.} \tag{20}$$

Representatives of the GDOD felt that the expected incentive profit should be positive so long as expected cost is below the threshold $\bar{E}$. The budget-based schemes will satisfy this requirement provided that $a(\bar{E}) = 0$ (and hence $\alpha[E] > 0$ for $E < \bar{E}$).[29] One consequence of this specification is that the firm will never be worse off in terms of expected profit than under a cost-plus-fixed-fee contract (or proportional-fee contract). It was emphasized that this "individual rationality" property is essential for widespread implementation of the budget-based schemes. Because of political pressure the GDOD would find it difficult to implement the schemes if defense contractors perceived them as an attempt to "squeeze" contractor profits. In contrast, the budget-based schemes become more palatable if introduced as a mechanism for improving efficiency and making both sides better off in the process.

Many cost-based government contracts are subject to price ceilings.[30] With a cost-plus-fixed-fee contract, the imposition of a price ceiling is effectively the only instrument the government has to contain cost overruns. Though the above model analysis suggests that there is no need for such ceilings given proper incentive provisions, representatives of the GDOD decided to maintain price ceilings in conjunction with the budget-based schemes. To ensure that the resulting project price never exceeds the price ceiling, say $\bar{P}$, actual cost can be reimbursed only up to some cost

---

[27]See LSP, No. 50–52. The law does not stipulate, though, that incentive profits have to be positive.

[28]The suggestion to interpret $a(E)$ as a reward for the firm's willingness to reveal its favorable cost information was rejected by one procurement officer with the argument: "By law, they are required to provide us that information anyhow, so we can't pay a premium for that!!"

[29]The boundary condition $a(\bar{E}) = \underline{0}$ may bias the firm's cost estimate if $E > \bar{E}$. For firms to report truthfully expected cost in excess of $\bar{E}$, it may be necessary to set $a(\bar{E}) < 0$.

[30]Cost-plus-incentive-fee contracts that are subject to a price ceiling are usually referred to as fixed-price-incentive contracts (see GAO, 1987).

ceiling K. If the firm issues the highest possible estimate $E = \overline{E}$, the value of K is obtained by solving the linear equation:

$$\overline{P} = K + BP(K) + a(\overline{E}) + b(\overline{E}) \bullet (\overline{E} - K), \tag{21}$$

where $BP(x)$ denotes the firm's base profit (as mentioned above, the current rule in Germany is roughly $BP(x) = 0.05 \bullet x$. For simplicity, it was decided further to set $K = \overline{E}$. Since $a(\overline{E}) = 0$, Eq. (21) then reduces to $\overline{P} = K + BP(K)$. For cost estimates below $\overline{E}$, it would be possible to increase the value of the cost ceiling without violating the overall price ceiling.[31] In the application described below, however, the government opted for one cost ceiling only.

The introduction of cost ceilings may distort the incentives created by the budget-based schemes. Since the contracting firm bears 100 percent of all costs in excess of the ceiling K, the compensation scheme is piece-wise linear with a kink at K. Recall that, for the model described in Section 2, the firm has partial control over the expected project cost $z = \mu + \alpha(\mu)$, yet actual cost x is a random variable with some probability distribution $F(x|z)$. For concreteness, suppose that $x = z + \varepsilon$, where ε denotes a random variable with mean zero and support $[\underline{\varepsilon}, \overline{\varepsilon}]$. If $z + \varepsilon > K$, and therefore actual cost x exceeds the ceiling K with positive probability, the firm's incentives will be affected in two ways. First, there will be a stronger incentive to reduce the amount of slack $\alpha(\mu)$. However, strengthening the firm's performance incentives will not be in the government's interest if the original budget-based scheme already provided optimal performance incentives.

Second, the presence of a cost ceiling affects the firm's reporting incentives. It is no longer in the firm's interest to provide an unbiased cost estimate (i.e., to report $E = z$). Instead, the firm will be better off by underestimating the expected cost (i.e., by reporting $E < z$). To see this, note that without cost ceilings the firm would not want to issue an unrealistically low cost estimate because cost overruns would be penalized at a higher rate. In the presence of cost ceilings, however, cost overruns in excess of K are penalized at the same rate (100 percent) irrespective of the cost estimate. Therefore, the firm will have an incentive to bias its cost estimate downward. The resulting distortions tend to be smaller as the noise term ε centers more of its probability weight around the mean zero. In general, though, there is no evidence that the budget-based schemes will remain optimal incentive mechanisms in the presence of price and cost ceilings.

Implementation for Pilot Projects

First applications of the budget-based schemes were undertaken in the fall of 1989 and spring of 1990 in the context of two development contracts (with two separate suppliers). Procurement officers of the GDOD indicated that the use of a budget-

---

[31] Let $K^*(E)$ denote the cost ceiling corresponding to the estimate E. The value of $K^*(E)$ solves the linear equation: $P = K^*(E) + BP[K^*(E)] + a(E) + b(E) \bullet [E - K^*(E)]$. Since $b(\bullet)$ is decreasing and $a'(E) = -b(E)$, it follows directly that $K^*(E)$ is decreasing in E. Furthermore, $K^* \overline{E} = \overline{E}$.

based scheme was decided fairly late during the contract negotiations in both cases. Originally, the government intended to conduct both projects as cost-plus-proportional-fee contracts (with a markup of approximately 5 percent) subject to a price ceiling. In fact, the government negotiators presented the particular budget-based schemes only after the price ceilings had been negotiated. Table 7.1 shows the matrix of incentive profits for one of the projects.

Table 7.1 shows that if the contracting firm were to issue the estimate $E = K$, its incentive profit would be zero if actual costs were to be greater than or equal to K. For actual cost x less than K, the firm would receive 15 percent of the budget variance $(K - x)$ as its incentive profit. By providing the estimate $E = K$, the contracting firm would always be better off than under the original cost-plus-proportional-fee contract. It would receive profits at least as high irrespective of actual cost.

The contract designers apparently felt that in the best of circumstances the contracting firm might issue a cost estimate of 70 percent of K. In accordance with the prescriptions in Sections 2 and 3, the cost-share parameters decrease monotonically over the interval (70, 100), and range from 0.5 to 0.15. The corresponding target profits $a(E_i)$ are the diagonal entries in the matrix. It can be verified that these numbers satisfy the recursive equations in (15) (Observation 3), with the boundary condition given by $a(100) = 0$ (the matrix contains rounding errors because the numbers have been computed only to the first decimal point). Under this scheme, the firm's total profit will always be positive. Even when the incentive profit results in a large penalty ($E = 70$ percent of K and $x = K$), the firm still earns a profit of $(0.05 - 0.044)K = 0.006K$.

The procurement officers who designed the budget-based scheme in Table 7.1 did not reveal how the cost-share parameters $b(E)$ were actually determined. In particular, it was not made public to what extent the designers of the scheme formally assessed a prior distribution $N(\mu)$ and a benefit function $B(\alpha)$. It should be noted, though, that the cost-share parameters chosen are consistent with the model and the constructive procedure described above. In particular, one can construct functions $B(\alpha)$ and $N(\mu)$ such that the resulting optimal cost-share parameters coincide with the ones in Table 7.1.

Table 7.1

Matrix of Incentive Profits for One GDOD Project

		Actual Cost (as a percentage of project cost ceiling)						Cost-Share Parameter
		70	75	80	85	90	95	100	(%)
Estimated cost	70	10.6	8.0	5.6	3.1	0.6	−1.9	−4.4	50
(as a percentage	75	10.3	8.1	5.8	3.6	1.3	−0.9	−3.2	45
of project	80	10.0	7.9	5.9	3.9	1.9	−0.2	−2.3	42
cost ceiling)	85	9.6	7.8	5.8	4.1	2.3	0.4	−1.5	37
	90	9.0	7.3	5.7	4.0	2.5	0.8	−0.8	32
	95	8.1	6.7	5.3	3.9	2.4	1.1	−0.3	26
	100	4.4	3.6	2.9	2.2	1.4	0.8	0.0	15

In response to the budget-based scheme shown in Table 7.1, the contracting firm submitted a cost estimate of 85 (i.e., $E = 85\% \cdot K$). As a consequence, it selected a cost-share parameter of $b = 0.37$, and the following total profit equation:

$$\text{Profit } (x) = 0.05 \cdot x + 0.041 \cdot K + 0.37(0.85 \cdot K - x) \tag{22}$$

for $x \leq K$. In case $x > K$, the firm's profit is "frozen" at K; that is, it is given by Eq. (22) evaluated at $x = K$. The actual cost figure will be known sometime in the near future upon completion of the project. It would then be interesting to find out in some detail how this incentive scheme affected the conduct of the project.

In general, one may ask whether the government should be confident *ex ante* that the above incentive scheme will perform better than the previous cost-plus-proportional-fee contract. Leaving aside the theoretical predictions developed in Section 2, one could argue that even without the above incentive provision, the firm's cost might remain below the ceiling K. If so, the budget-based scheme would result only in additional profits for the firm. However, such an argument would overlook the empirical finding that in the past the actual price paid by the German government was within 5 percent of the price ceiling for more than 90 percent of all cost-plus-proportional-fee contracts.

5. CONCLUDING REMARKS

Although this article has focused on procurement contracting, there are some indications that budget-based schemes might also be useful in the context of public utility regulation. As in procurement contracting, a major issue in utility regulation is that the firm has better information regarding production costs and that many of its actions are unobservable and subject to moral hazard. The predominant method for regulating public utilities in the United States (and many other countries) is rate-of-return regulation. As observed by numerous practitioners and academic researchers, a major drawback of rate-of-return regulation is that the utility has little incentive to reduce or contain its operating costs. It is therefore not surprising that regulators have experimented with profit-sharing arrangements similar in spirit to the cost-plus-incentive-fee contracts discussed in Section 2. Brown et al. (1989) discuss the types of incentive plans that have been adopted by state regulators.[32]

In testimony before the Federal Energy Regulatory Commission, Brown (1990) advocates the use of a budget-based scheme for adjusting the rate base of a power plant in Michigan.[33] The perceived benefits of such an incentive scheme include unbiased estimation of future capital spending by the utility as well as more reliable estimates

---

[32]A prominent example of profit sharing is the so-called "sliding scale" of rates adopted for the Potomac Electric Power Company. Bussing (1936) reports that in 1925 the Public Utilities Commission of the District of Columbia granted the utility a basic 7.5 percent rate of return on its rate base. If in any particular year actual profit exceeded the target 7.5 percent return, half of the excess profit was to be absorbed the next year. Accordingly, if all other variables remained constant, the utility's output prices were to be lowered the next year, so that operating profit would be reduced by half of the previous year's excess profit.

[33]The rate base comprises those long-term assets on which the firm is allowed to earn its rate of return.

of the resulting electricity prices. Perhaps more important, the utility would have an incentive to reduce its capital spending, depending on the magnitude of the cost-share parameter chosen. Of course, this incentive would even be stronger under a price cap arrangement, yet, as discussed before, fixed-price contracts are likely to result in larger informational rents that would have to be paid for by consumers.

APPENDIX

It is shown in Kirby et al. (1991) that the optimal $\alpha(\mu)$ is given by point-wise minimization of the government's expected total cost (project cost plus profit paid to the firm):

$$(\mu + \alpha) - B(\alpha) + \frac{N(\mu)}{n(\mu)} \bullet B'(\alpha) , \qquad (A.1)$$

where $\alpha \geq 0$. Under the assumptions of the lemma, the resulting function $\alpha(\mu)$ will be (weakly) increasing.

To verify Observation 1, we note that if

$$\frac{N(\mu)}{n(\underline{\mu})} = 0,$$

then Eq. (A.1) is minimized at $\alpha = 0$ for $\mu = \underline{\mu}$. This, in turn, implies $b(\underline{\mu}) = b(\underline{E}) = B'(0).$[34] If

$$\frac{N(\mu)}{n(\underline{\mu})} > 0,$$

then $\alpha(\mu) \geq 0$, which implies that $b(\underline{E}) \geq B'(0)$. To see that $b(\underline{E}) \geq v_2$ (where $v_2 = B[\alpha^{-*}]$), let μ^* be the lowest value of μ such that $\alpha(\mu) = \alpha^*$.[35] Since $B'(\alpha) = v_2$ for $\alpha \geq \alpha^*$, the expression in Eq. (A.1) is minimized by $\alpha(\mu) = \alpha^*$ for all $\mu \geq \mu^*$. Hence $b(E) = v_2$ for all $E \geq \mu^* + \alpha^*$.

To prove Observation 2, we substitute the linear-quadratic function $B(\alpha)$ (as given in Eq. (12) into Eq. (A.1) and solve for $\alpha(\mu)$:

$$\alpha(\mu) = \begin{cases} 0 & \text{for } \mu \leq \mu_*, \\ \dfrac{s_1 - 1}{s_2} + \dfrac{N(\mu)}{n(\mu)} & \text{for } \mu_* \leq \mu \leq \mu^*, \text{ and} \\ \alpha^* & \text{for } \mu^* \leq \mu. \end{cases} \qquad (A.2)$$

---

[34]Obviously, the government cannot benefit by setting $b > B'(0)$.

[35]Depending on the value of $\overline{E}$, it may be that $\alpha(\mu) < \alpha^*$ for all μ satisfying $\mu + \alpha(\mu) \leq \overline{E}$. If so, $b(E) > v_2$ for all admissible cost estimates.

The values μ_* and μ^* are determined by the equations:

$$\frac{1 - s_1}{s_2} = \frac{N(\mu_*)}{n(\mu_*)} \text{ and } \frac{s_1 - 1}{s_2} + \frac{N(\mu^*)}{n(\mu^*)} = \alpha^*.$$

Recalling Eq. (9), we obtain:

$$b\big[\mu + \alpha(\mu)\big] = s_1 - s_2 \bullet \alpha(\mu), \tag{A.3}$$

if $\mu \leq \mu^*$, and $b(\mu + \alpha^*) = s_1 - s_2 \bullet \alpha^*$ for $\mu > \mu^*$. It will be convenient to rewrite Eq. (A.3) as:

$$b(z) = s_1 - s_2 \bullet \big[z - \gamma(z)\big], \tag{A.4}$$

where $\mu^* + \alpha^* \geq z \geq \mu_* + \alpha(\mu_*)$, and $\gamma(z)$ is the inverse of the map $\phi(\mu)$ defined by $\phi(\mu) \equiv \mu + \alpha(\mu)$. Differentiation of Eq. (A.4) yields:

$$b'(z) = -s_2 \bullet \big[1 - \gamma'(z)\big].$$

Therefore, $b'(z)$ will be increasing (equivalently $b[z]$ will be convex) if $\gamma(z)$ is convex. This will be the case if $\phi(\mu)$ is concave, or, in light of Eq. (A.2), if the function $h(\mu)$ is concave. This completes the proof of Observation 2.

The following is an explicit characterization of the optimal cost-share parameters for the special case where the benefit function $B(\alpha)$ and the prior distribution $N(\mu)$ satisfy Eqs. (12) and (13), respectively.

Claim: Given Eqs. (13) and (14), the optimal cost-share parameters $b(E)$ are given by:

$$b(E) = \begin{cases} s_1 & \text{for } E_0 \leq E \leq z, \\ s_1 - s_2 \bullet \left[\dfrac{1}{r+2}(E - z_1)\right] & \text{for } z_1 \leq E \leq z_2, \\ s_1 - s_2 \bullet \tau(E) & \text{for } z_2 \leq E \leq z_3, \text{ and} \\ s_1 - s_2 \bullet \alpha^* & \text{for } z_3 \leq E \leq \bar{E}, \end{cases} \tag{A.5}$$

where:

$$z_1 = \underline{\mu} + (r+1) \bullet \frac{1-s_1}{s_2},$$

$$z_2 = m + \frac{1}{r+1}(m - z_1),$$

$z_3 = y + \alpha^*$, with y solving the equation,

$$\alpha^* = \frac{1}{r+1}\left[2 \bullet \frac{(\bar{\mu} - m)^{r+1}}{(\bar{\mu} - y)^r} - (\bar{\mu} - y) - (z_1 - \underline{\mu})\right].$$

The function $\tau(E)$ is increasing and convex with

$$\tau(z_3) = \alpha^*, \tau(z_2) = \frac{1}{r+2}(z_2 - z_1),$$

$$\tau'(E) \geq \frac{1}{r+2} \quad \text{for all } E \in (z_2, z_3).$$

Proof: For the probability distributions in Eq. (13) we find that

$$\frac{N(\mu)}{n(\mu)} = \begin{cases} \dfrac{1}{r+1}(\mu - \underline{\mu}) & \text{if } \mu \leq m \text{ and} \\[2ex] \dfrac{1}{r+1}\left[2 \bullet \dfrac{(m-\underline{\mu})^{r+1}}{(\bar{\mu} - \mu)^r} - (\bar{\mu} - \mu)\right] & \text{if } \mu \geq m. \end{cases} \qquad (A.6)$$

Combining Eqs. (A.2) and (A.6), we obtain:

$$\alpha(\mu) = \begin{cases} 0 & \text{if } \mu \leq \mu*, \\[1.5ex] \dfrac{1}{r+1}(\mu - \mu*) & \text{if } \mu* \leq \mu \leq \underline{\mu}, \\[1.5ex] \dfrac{1}{r+1}\left[2 \bullet \dfrac{(m-\underline{\mu})^{r+1}}{(\bar{\mu} - \mu)^r} - (\bar{\mu} - \mu) - (\mu*-\underline{\mu})\right] & \text{if } m \leq \mu \leq \mu^*, \text{and} \\[1.5ex] \alpha^* & \text{if } \mu \geq \mu^*. \end{cases}$$

Substitution in the above equations for $\mu*$ and μ^* implies that:

$$\mu* = \underline{\mu} + (r+1) \bullet \frac{1-s_1}{s_2},$$

and μ^* is given as the solution to the equation:

$$\alpha^* = \frac{1}{r+1}\left[2 \bullet \frac{(m-\underline{\mu})^{r+1}}{(\bar{\mu} - \mu^*)^r} - (\bar{\mu} - \mu^*) - (\mu*-\underline{\mu})\right].$$

Correspondingly, we obtain the cost estimates $z_i = \mu_i + \alpha(\mu_i)$, with $z_2 = m + \alpha(m)$, as given in the statement of the lemma.

The lowest cost-share parameter inducing $\alpha(\mu) = 0$ for $\mu_* > \mu > \underline{\mu}$ is $B'(0) = s_1$. Thus, $b(E) = s_1$ for $E_0 = \underline{\mu} \le E \le \mu_* = z_1$. For $\mu^* > \mu > \mu_2$, the firm's choice of slack is interior, satisfying

$$b[\mu + \alpha(\mu)] = s_1 - s_2 \bullet \alpha(\mu),$$

or, equivalently,

$$b(E) = s_1 - s_2 [E - \phi^{-1}(E)],$$

with $\mu + \alpha(\mu) \equiv \phi(\mu) = E$. Since $\phi(m) = z_2$, it follows that

$$\phi(\mu) = \mu + \frac{1}{r+1}(\mu - \mu_*),$$

for $\mu_* < \mu < m$, and thus,

$$b(E) = s_1 - s_2 \left[\frac{1}{r+2} \bullet (E - z_1) \right]$$

for $z_2 \ge E \ge z_1$. Finally, $z_3 = \mu^* + \alpha(\mu^*) = \mu^* + \alpha^*$. It follows immediately that $\alpha(\mu) = \alpha^*$, for $\mu \ge \mu^*$ provided $b(E) = s_1 - s_2 \bullet \alpha^*$ for $E > z_3$.

The claims regarding the function $\tau(\bullet)$ remain to be proved. Convexity of the function $\tau(\bullet)$ follows from Observation 2. Continuity of the function $\alpha(\bullet)$ combined with the fact that $\tau(E) = E - \phi^{-1}(E)$ implies that

$$\tau(z_2) = \frac{1}{r+2}(z_2 - z_1)$$

and $\tau(z_3) = \alpha^*$. Differentiation of $\tau(\bullet)$ gives,

$$\tau'(E) = 1 - \frac{1}{\phi'[\phi^{-1}(E)]},$$

or, alternatively,

$$\tau'[\phi(\mu)] = 1 - \frac{1}{1+\alpha'(\mu)} = \frac{\alpha'(\mu)}{\alpha'(\mu)+1}. \tag{A.7}$$

Thus, $\tau(\bullet)$ is monotone increasing. Differentiating $\alpha(\mu)$, as given by [ii] on the interval (m, μ^*) yields:

$$\alpha'(\mu) = \frac{1}{r+1} \left[2r \bullet \left(\frac{\bar{\mu} - m}{\bar{\mu} - \mu} \right)^{r+1} + 1 \right].$$

Therefore,

$$\alpha'(\mu) \geq \frac{1}{r+1},$$

and:

$$\frac{\alpha'(\mu)}{1 + \alpha'(\mu)} \geq \frac{1}{r+2},$$

proving that

$$\tau'(E) \geq \frac{1}{r+2},$$

for $z_3 > E > z_2$.

PROCUREMENT POLICY AND CONTRACTING EFFICIENCY
by Anthony G. Bower[1]

1. INTRODUCTION

This paper examines contractor incentives in a two-stage model of procurement that incorporates adverse selection, moral hazard, and discretionary learning. Optimal contracts are characterized for a situation in which the government can audit the winning firm's costs at specified times during the engagement. The model is applicable to contracts both let through negotiations with a single supplier and those let through competitive bidding. While the model is general enough to include many procurement situations with cost observability, it will be especially applied to defense contracting, in which the multistage framework allows for analysis of such real-world features as multiple audits, partial commitment to the terms of the contract, and learning by doing.

The motivations are two: first, the theory of regulation and optimal contracting under incomplete information; and second and perhaps most important, from the study of the *relative value* of various procurement contract options that a principal has at its disposal. Regarding the first, for the case of a fixed procurement quantity, the analysis extends the theory of contracts under incomplete information. Similar to Riley (1988), I allow the principal to observe *ex post* a signal that is correlated with the agent's private information (i.e., the government may audit the firm's costs). The model is a generalization to multiple periods of the one-shot models of Laffont and Tirole (1986, 1987) and McAfee and McMillan (1986). The generalization allows examination of several important issues in contracting that cannot be addressed in the simpler, one-period frameworks. These issues include the value of multiple audits and the size of the loss from imperfect commitment to the terms of the contract.[2]

[2]The model is also more descriptive of actual DoD practice and is designed to address specific DoD contracting problems. It is related to, and addresses some of the same issues as, Baron and Besanko (1984, 1987a) and Riordan and Sappington (1987, 1988). Baron and Besanko study a regulatory relationship under regimes of limited commitment and "fairness." Riordan and Sappington examine the design of procurement contracts under limited commitment and the related problem of awarding monopoly franchises.

Like Laffont and Tirole, who use a one-shot model, I show for a multiperiod model that a menu of linear incentive contracts can be optimal.

The optimal contract from an *ex ante* perspective requires the government to commit not to revise contracts even though both the contractor and the government could be made better off *ex post* by doing so.[3] If this commitment is not credible, the government must pay more to obtain truthful reports. The model provides a bound on the additional procurement cost that results from the renegotiation.

The second motivation is policy relevance: The analysis allows for substantial flexibility in evaluating the relative worth of alternative procurement instruments available to the principal. This paper analyzes four procurement instruments, which the theoretical contracting literature has identified as optimal to use if the costs of using these instruments is zero. In this paper, the analysis includes all four instruments at once, but the conclusion is no different: If implementation costs are zero, then all four options are useful. However, in the real world, all of the procurement instruments available cost the principal something to employ and unlike most previous papers in contracting, this paper identifies the relative benefits of each of the instruments. Those benefits can then be weighed against the potential costs to serve as guidelines in determining the optimal contract under nonzero implementation costs. For an example of how tangible these implementation costs are, consider defense contracting. The institutional literature of defense contracting (for example, Fox, 1974, Peck and Scherer, 1962, and Gansler, 1989) emphasizes that regulating and administering contracts consumes enormous resources. Gansler notes that 27,000 people are employed by DoD in administering contracts. The use of a menu requires more detailed and complex analysis on the part of the government and the contractor. Optimal contracts that induce Pareto inefficiencies in later periods may be difficult for DoD to defend against budget cutters. The use of audits requires a large salaried team of auditors, and inducing competition incurs search and bid preparation costs.

Since administering contracts is costly, the question arises as to which procurement instrument delivers the most value. This paper provides an answer for some special cases. First, the procurement cost when a menu of contracts is used is compared to the procurement cost when a single incentive contract is used. In numerical examples for a particular specification, the difference in cost between the two alternatives is found to be always less than 2 percent of total cost. Also, perhaps counterintuitively, it is found that the value of a menu may be *higher* with more firms bidding. That is , a menu's value may not be highest for the monopoly case. Second, the procurement cost of contracts that use zero, one, or two audits is compared. In the model, audits are assumed to be costless. In the examples, the value of auditing is found to be large for one audit but to increase very little for more frequent audits. Third, the expected cost of limited commitment (defined as commitment plus *ex post*

---

[3]This requirement is particularly relevant in studying defense contracting, because contracts are often very long and for a variety of reasons are frequently renegotiated. Gansler (1989) points out that the lag from exploratory development to initial deployment of a weapons system is 12 to 15 years and has been lengthening. This is one reason for renegotiation.

renegotiation with the initial contract as the status quo), compared to full commitment, is shown often to be a fairly small percentage of total expenditures. Fourth, the gains to increasing the number of bidders, each of which draws from an identical cost distribution, are relatively much larger. Of course, all types of contracts tend to perform equally well in the limit as the number of bidders increases, and converge to the first-best, fixed-price contract as the number of bidders increases to infinity (i.e., the optimal contract in the limit is fixed price). Overall, one ramification of the analysis is that the general form of the contracts used by DoD is "relatively" good, in the sense that those restricted contract forms can achieve nearly optimal results.

In Section 2, the model is introduced and solved for the general cases of commitment and limited commitment, given auditing of total cost. In Section 3, numerical solutions are obtained assuming a uniform distribution of costs and quadratic cost of effort. In Section 4, the main results are summarized.

2. THE MODEL

2.A. Full Commitment

2.A.1. Basic model. A government wishes to maximize social welfare. A good of given quality and quantity is procured from a single profit-maximizing firm and the government's demand is perfectly inelastic.[4] Both quantity and total production time are normalized at one. Assume for now that the government can credibly commit to any incentive scheme. The issue of commitment is important and is taken up in part B of this section.[5]

Index time as t. The time line is shown in Figure 8.1. At time zero ($t = 0$), the firm communicates with the government and then begins to produce. The period before time λ is the first period, and after time λ is the second period. At time λ and at time 1 the government has the opportunity to audit the firm and to base reimbursement on the audits. The auditing time λ is exogenous and may be considered the

RANDMR462-8.1

	First audit	Second audit, end of contract
Communication		

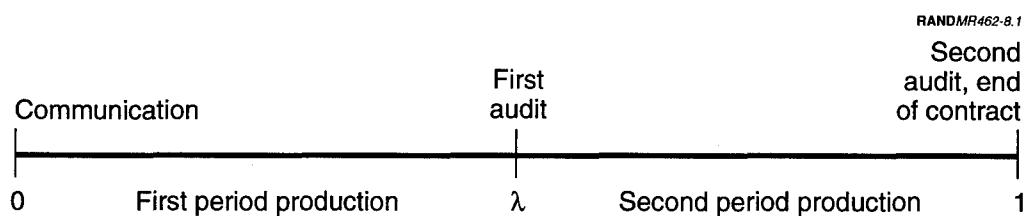| 0 | First period production | λ | Second period production | 1 |

Figure 8.1—Timeline

---

[4]The generalization to multiple firms is straightforward and will be addressed in subsection 2.A.7.

[5]In general, agencies of the government have difficulties committing to policies, particularly if the relationship spans many years. For an exploration of these issues, see Baron and Besanko (1987a).

earliest feasible audit time.[6] The government can audit the firm's costs perfectly and costlessly.[7] The government is assumed to observe *total* (or average) cost.

The initial cost C of production for the firm is distributed according to a common knowledge distribution function $F(\bullet)$ with density $f(\bullet)$ on a closed interval $[C_{min}, C_{max}]$, with the usual assumption that $F(C)/f(C)$ is increasing.[8] The firm's cost is its private information.

The firm may reduce costs by exerting unobservable, costly effort. Effort permanently improves production techniques; such improvements are referred to as discretionary learning. For simplicity, the firm may set effort only twice: It sets first-period effort η_1 at time 0, and second-period effort η_2 at time $\lambda, \lambda \in (0,1)$.[9] Exerting effort η_i for a length of time dt reduces costs permanently by $\eta_i \bullet dt$ at a cost of $\psi(\eta_i) \bullet dt$.[10] The cost $\psi(\eta_i) \bullet dt$ is borne by the firm and cannot be reimbursed directly. The assumption that the firm bears the full burden of the cost-deduction effort is made without loss of generality, for any cost reduction that can be reimbursed directly may be imbedded in C. It is assumed that $\psi(\bullet)$ is strictly increasing and convex. In contrast to Laffont and Tirole (1986), in which marginal cost drops immediately to the terminal value, in this learning model the rate of cost flow declines gradually over time, to $C - \lambda\eta_1 - (1-\lambda)\eta_2$ at time 1. Total costs C_i in period i are $C_1 = \int_0^\lambda (C - \eta_1 t)dt = \lambda C - (\lambda^2/2)\eta_1$ and $C_2 = \int_\lambda^1 (C - \lambda\eta_1 - \eta_2(t-\lambda))dt = (1-\lambda)C - \lambda(1-\lambda)\eta_1 - [(1-\lambda)^2/2]\eta_2$. The government may observe C_1 at time λ and C_2 at time 1 perfectly by auditing the firm. Finally, the interest rate is assumed to be zero.[11]

This is a noncooperative game with asymmetric information. Without loss of generality, the game can be represented by a direct revelation mechanism, and the revelation principle implies that attention may be restricted to mechanisms in which the firm truthfully reports its type (initial cost) at $t = 0$. To induce the firm to report truthfully, the mechanism must be incentive compatible (IC). It also must be individually rational (IR) for the firm to participate. Define $\hat{C}$ as the firm's cost report. A mechanism M is a transfer function $R(\bullet)$, which specifies the payment and cost targets for the firm as a function of the report and audited costs, or $M = (R(\hat{C}, C_1, C_2), C_1(\hat{C}), C_2(\hat{C}))$. The government announces M before the firm's report. The payment $R(\bullet)$ is assumed to be made at $t = 1$ and induces effort η_1^* and η_2^* from the firm. Deriving the optimal mechanism is an exercise in optimal control theory. It is convenient to use η_1 and η_2 (which the government cannot observe, but may infer) as the control variables.

---

[6]Thus, the model allows for auditing periods of different lengths. The numerical examples presented below will consider the special case of $\lambda = 1/2$.

[7]The restriction of perfect auditing will be loosened in subsection 2.A.6.

[8]A more general distribution could be assumed but would complicate the analysis and exposition substantially.

[9]None of the qualitative results presented in this section is affected by this assumption.

[10]Alternatively, effort can be thought of as the opposite of unobservable "waste" or "shirking." High effort, for example, corresponds to low waste. The mathematics of the model can adopt either interpretation.

[11]See Bower and Osband (Chapter Four) for an application with nonzero interest rates.

The remainder of this subsection deals with the full-commitment case, and is organized as follows. I start with the simplest case of one firm and deterministic costs. In part A.2, the profits required to induce the firm to participate and truthfully reveal its cost are derived. Using these requirements, the optimal transfer function $R(\bullet)$ for a monopolist with no randomness in costs is derived in A.4. (As a yardstick to compare the second-best solution, the first-best solution is provided in A.3.) The solution is interpreted in the context of DoD profit policy in A.5. The solution is then generalized to the case of uncertainty in costs in A.6, and generalized further to multiple firms with uncertain costs in A.7. Subsection B performs the corresponding analysis (in abbreviated form) for the case of commitment plus renegotiation (called limited commitment).

2.A.2. Individual rationality and incentive compatibility conditions. In this subsection, assume that there is only one firm and deterministic costs. The incentive compatibility condition for the full-commitment problem (ICF), which quantifies the rents to the firm's private information, is derived below.

The firm reports a cost $\hat{C}$. The mechanism specifies a payment function $R(\hat{C}, C_1(\hat{C}), C_2(\hat{C}))$ and cost targets $\hat{C}_1(\hat{C}), \hat{C}_2(\hat{C})$. It is assumed that the firm must meet the targets; i.e., if a firm does not meet its targets, it suffers an infinite penalty (later it will be verified that this restriction can be lifted without changing the solution). Define $\eta_1(\hat{C}|C)$ as the firm's effort given its report $\hat{C}$ and its true cost C. Define first-period effort $\eta_1(C)$ given a truthful report C as $\eta_1(C) \equiv \eta_1(C|C)$. For now, assume that the government has perfect observability of costs *ex post*. Thus, the firm will choose $\eta_1(\hat{C}|C)$ such that actual cost C_1 equals targeted costs $\hat{C}_1$, or

$$\hat{C}_1(\hat{C}) \equiv \lambda\hat{C} - \frac{\lambda^2}{2}\eta_1(\hat{C}) = C_1(C, \hat{C}) \equiv \lambda C - \frac{\lambda^2}{2}\eta_1(\hat{C}|C),$$

so that

$$\eta_1(\hat{C}|C) = \eta_1(\hat{C}) + \frac{2}{\lambda}(C - \hat{C}). \tag{1}$$

Observe in Eq. (1), that for $\hat{C}$ large enough, $\eta_1(\hat{C}|C)$ is negative. However, it is assumed that $\eta_i \geq 0$ and so any report that implies $\eta_i(\hat{C}|C) < 0$ will be treated by the government as a report that implies $\eta_i(\hat{C}|C) = 0$.

At the time of the second audit, audited cost must again equal targeted cost:

$$\hat{C}_2(\hat{C}) \equiv (1-\lambda)\hat{C} - \lambda(1-\lambda)\eta_1(\hat{C}) - \frac{(1-\lambda)^2}{2}\eta_2(\hat{C})$$

$$= C_2(C, \hat{C}) \equiv (1-\lambda)C - \lambda(1-\lambda)\eta_1(\hat{C}|C) - \frac{(1-\lambda)^2}{2}\eta_2(\hat{C}|C)$$

or

$$\eta_2(\hat{C}|C) = \eta_2(\hat{C}) + \frac{2}{1-\lambda}(C - \hat{C}) + \frac{2\lambda}{1-\lambda}(\eta_1(\hat{C}) - \eta_1(\hat{C}|C)). \tag{2}$$

Substituting Eq. (1) yields

$$\eta_2\,(\hat{C}\mid C) = \eta_2\,(\hat{C}) - \frac{2}{1-\lambda}\,(C - \hat{C}).$$ (3)

Equations (1) and (3) identify the efforts required to meet the cost targets for a report $\hat{C}$ when the true cost is C.

The firm's net profit equals reimbursements from the government less its costs of production less its cost of effort. Reimbursement is, without loss of generality, a lump-sum transfer $S(\hat{C})$ plus observed cost:

$$R(\hat{C}, C_1, C_2) = S(\hat{C}) + C_1 + C_2.$$

The payment R is made if and only if $C_1 = \hat{C}_1$ and $C_2 = \hat{C}_2$; otherwise, the firm suffers an infinite penalty. Thus, profit $U(\hat{C}\mid C)$ of a firm with cost C announcing $\hat{C}$ is

$$U(\hat{C}\mid C) = S(\hat{C}) - \lambda\psi\,(\eta_1(\hat{C}\mid C)) - (1-\lambda)\psi(\eta_2(\hat{C}\mid C)).$$ (4)

Substituting Eqs. (1) and (3) into Eq. (4) gives

$$\begin{aligned}
U(\hat{C}\mid C) = {} & S(\hat{C}) - \lambda\psi\!\left(\eta_1(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right) \\
& - (1-\lambda)\psi\!\left(\eta_2(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right).
\end{aligned}$$ (5)

Defining $U(C) \equiv U(C\mid C)$, and using the envelope theorem,[12] yields the local incentive compatibility condition

$$U'(C) = -2\psi'(\eta_1(C)) + 2\psi'(\eta_2(C)).$$ (ICF)

The following is an explanation of the intuition for the ICF constraint and of why it differs from the IC constraint in Laffont and Tirole (1986).[13] In the basic Laffont-Tirole model, if a firm marginally exaggerates its cost parameter ($\hat{C} = C + dC$), it receives a higher cost target. But since its productivity is higher than it claims it is, it can meet this higher cost target with an effort level that is lower than $\eta(\hat{C})$ on which its payment is based. The gain to the firm from overstating its cost parameter is equal to the difference between the effort costs it is compensated for and the effort costs it actually must bear to meet the cost target. To a first order of approximation, this gain is equal to $\psi'(\eta(C))\ dC$.

---

[12]It is straightforward but tedious to show that the derivative $\partial\hat{C}/\partial C$ (needed for the application of the envelope theorem) exists almost everywhere.

[13]I am grateful to a referee for suggesting this intuition and the formulation presented in the next footnote.

Now, this same effect operates in this model, insofar as the first-period effort level η_1 is concerned. This explains the first term in (ICF) (except for the factor of two, which reflects the fact that effort affects costs gradually, through learning). However, in this model, first-period and second-period effort are substitutes with respect to second-period production costs. If the firm exaggerates its cost parameter ($\hat{C} = C + dC$) and reduces first-period effort, it is forced to increase second-period effort to compensate for the reduced first-period effort. Indeed, the second-period effort level $\eta_2(\hat{C}|C)$ that must be expended to meet $\hat{C}_2$ is actually greater than the effort $\eta_2(\hat{C})$ on which its payment is based (see Figure 8.2). Thus, there is a *cost* of misreporting that, to a first-order approximation, is proportional to $2\psi'(\eta_2(C))$ dC. This "effort-substitution" effect explains the second term in (ICF) and works against the usual rent-creation effect.[14]

The derivation of the IR constraint is substantially less work. The IR constraint is that profits are nonnegative for whatever cost the firm has. Thus $U(C) \geq 0$, $\forall C$.

2.A.3. First-best solution. Before proceeding to the solution with full commitment, it is informative for comparative purposes to identify the first-best solution, which is attained if the government can observe effort as well as cost. In this case, $U(C) = 0$ for
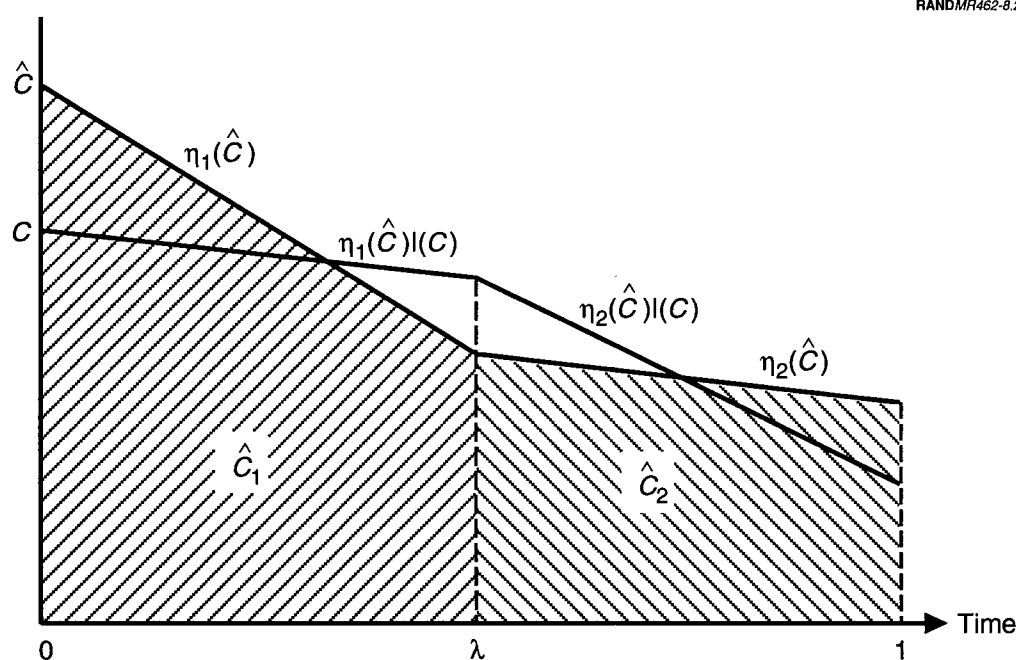

RAND*MR462-8.2*

Figure 8.2—Costs

---

[14]If the linkage between first-period and second-period cost is weakened, then this effort-substitution effect is also weakened. Imagine that the effects of first-period effort "depreciate," so that $C_2 = (1 - \lambda) C - \phi\lambda(1 - \lambda)\eta_1 - 1/2(1 - \lambda)^2 \eta_2$, where $\phi \in [0,1]$ is a depreciation parameter. Then (ICF) becomes $U'(C) = -2\psi'(\eta_1(C)) - 2[1 - 2\phi]\psi'(\eta_2(C))$, which implies that if $\phi < 1/2$, then rents will increase in both first- and second-period effort and the optimal contract will distort both downward. Note that if $\phi = 0$, the classic repeated solution of the one-shot game is obtained, because there is no linkage between periods.

all C, and the government pays the firm its costs, which include the cost reductions from first-best effort, so reimbursement $R^f(C)$ is

$$R^f(C) = C - \frac{\lambda^2}{2} \eta_1^f(C) - \lambda(1-\lambda)\eta_1^f(C) - \frac{(1-\lambda)^2}{2} \eta_2^f(C)$$
$$+ \lambda\psi(\eta_1^f(C)) + (1-\lambda)\psi(\eta_2^f(C)),$$

where the first-best effort levels are denoted η_1^f and η_2^f, respectively. Differentiate R^f with respect to η_1^f and η_2^f to obtain

$$\psi'(\eta_1^f(C)) = \frac{2-\lambda}{2}$$

$$\psi'(\eta_2^f(C)) = \frac{1-\lambda}{2}. \tag{FB}$$

First-period effort is greater than second-period effort because its benefits accrue in both periods. The results are summarized in Proposition 1.

Proposition 1: For the first-best case, for all C, effort is given in (FB) above. The firm makes zero profits, and the government pays the firm its cost C less net cost reduction from first-best effort.

2.A.4. Second-best solution with full commitment and one firm. The government realizes welfare W upon completion of the project. Assume that the social cost per dollar of transfers is $\alpha > 0$, which results because transfers are raised through some distortionary measure such as a tax.[15]

The government wishes to maximize W less total social cost and, since W is fixed, this problem is identical to minimizing total social cost. Using the local and global incentive compatibility condition, the second-best program with full commitment is

$$\min \int_{C_{min}}^{C_{max}} \left\{ (1+\alpha)\left[\lambda\psi(\eta_1(C)) - \frac{\lambda^2}{2} \eta_1(C) - \lambda(1-\lambda)\eta_1(C) \right.\right.$$

$$\left.\left. + (1-\lambda)\psi(\eta_2(C)) - \frac{(1-\lambda)^2}{2} \eta_2(C) + C \right] + \alpha U(C) \right\} f(C) \ dC \tag{P1}$$

$$s.t. \quad U'(C) = -2\psi'(\eta_1(C)) + 2\psi'(\eta_2(C)) \tag{ICF}$$

---

[15]This is identical to the Laffont and Tirole formulation. The parameter α may also measure the degree of dislike for firms' profits on the part of the regulator. Baron and Besanko have an interpretation in which a regulator "weights" a dollar of consumer surplus at 1 and a dollar of profits at a rate between 0 and 1. The Baron-Besanko weight on profits in this framework would equal $1/(1 + \alpha)$.

$$U(C) \geq U(\hat{C}|C), \qquad \forall \hat{C} \tag{GICF}$$

$$U(C) \geq 0. \tag{IR}$$

The second term inside the brackets is the savings $\int_0^\lambda \eta_1 \bullet t \; dt$ in the first period due to first-period effort. The third term is the savings in the second period due to first-period effort—this is the effect of learning. The first and fourth terms represent the disutility of effort. The fifth and sixth terms are the cost reduction from second-period effort and the cost of the firm, respectively. For now, the global incentive-compatibility condition (GICF) will be ignored. Global incentive compatibility will be checked in the proof of Proposition 3.

The second-best program P1 is an optimal control program with two controls. The Hamiltonian for the program is

$$H = \left\{ (1 + \alpha) \left[\lambda \psi(\eta_1(C)) - \frac{\lambda^2}{2} \eta_1(C) - \lambda(1 - \lambda)\eta_1(C) \right. \right.$$
$$\left. + (1 - \lambda)\psi(\eta_2(C)) - \frac{(1-\lambda)^2}{2}\eta_2(C) + C \right] + \alpha U(C) \left. \right\} f(C)$$
$$+ \mu(C)[-2\psi'(\eta_1(C)) + 2\psi'(\eta_2(C))] - \gamma(C)U(C).$$

The effort levels η_1 and η_2 are the controls, the profit $U(C)$ is the state variable, $\mu(C)$ is the co-state variable associated with the (ICF) condition, and $\gamma(C)$ is the nonnegative multiplier on the (IR) constraint. The necessary conditions on a solution, assuming that $\eta_1^*(C) > 0$ and $\eta_2^*(C) > 0$, are

$$f(C)(1+\alpha)\left(\lambda\psi'(\eta_1^*(C)) - \frac{\lambda^2}{2} - \lambda(1-\lambda) \right) - 2\mu(C)\psi''(\eta_1^*(C)) = 0 \tag{6a}$$

$$f(C)(1+\alpha)\left((1-\lambda)\psi'(\eta_2^*(C)) - \frac{(1-\lambda)^2}{2} \right) + 2\mu(C)\psi''(\eta_2^*(C)) = 0 \tag{6b}$$

$$\mu'(C) = -\frac{\partial H}{\partial U} = -[\alpha f(C) - \gamma(C)] \tag{6c}$$

$$U(C) \geq 0, \quad \gamma(C) \geq 0, \quad \gamma(C) \bullet U(C) = 0. \tag{6d}$$

Integrating Eq. (6c) and imposing the transversality condition $\mu(C_{min}) = 0$ yields

$$\mu(C) = -\alpha F(C) + \int_{C_{min}}^C \gamma(\upsilon) \; d\upsilon. \tag{7}$$

Rearranging Eq. (6a) establishes that

$$\psi'(\eta_1^*(C)) = \frac{2-\lambda}{2} - \frac{1}{1+\alpha}\frac{2}{\lambda}\frac{1}{f(C)}\psi''(\eta_1^*(C))\left(\alpha F(C) - \int_{C_{min}}^{C}\gamma(\upsilon)\ d\upsilon\right). \qquad (8)$$

Similar manipulations of Eq. (6b) establish that

$$\psi'(\eta_2^*(C)) = \frac{1-\lambda}{2} + \frac{1}{1+\alpha}\frac{2}{1-\lambda}\frac{1}{f(C)}\psi''(\eta_2^*(C))\left(\alpha F(C) - \int_{C_{min}}^{C}\gamma(\upsilon)\ d\upsilon\right). \qquad (9)$$

There are two possible types of solutions; a complete characterization of the solution for each type is provided below. In the first case, $U'(C) < 0$ (called the "unconstrained" case) for all values of C. This will occur if $C_{max} - C_{min}$ is not "too large." In this case, $U(C) > 0$ for all $C < C_{max}$. The second case (the "constrained" case) is when $U'(C) = 0$ for at least some portion of the cost distribution. A special case of this solution—a uniform cost distribution and quadratic $\psi(\bullet)$ (hereafter referred to as the "uniform-quadratic case")—is then used in the numerical examples in Section 3.

2.A.4.1. The unconstrained case. The following proposition characterizes the optimal solution for the case in which $U'(C) < 0$ for all C and $\psi(\bullet)$ is quadratic, where $\psi(\eta_i) = (K/2)\eta_i^2$ for some constant K.

Proposition 2: Assume that $\psi(\bullet)$ is quadratic and define $H(C) \equiv F(C)/f(C)$. Then, if $H(C_{max}) = 1/f(C_{max}) < [(1+\alpha)/\alpha][\lambda(1-\lambda)/4K]$, the unconstrained case holds and the optimal solution is characterized by:

(i) $\eta_1^*(C) = [(2-\lambda)/2K] - [\alpha/(1+\alpha)](2/\lambda)H(C)$. First-period effort is decreasing in C.

(ii) $\eta_2^*(C) = [(1-\lambda)/2K] + [\alpha/(1+\alpha)][2/(1-\lambda)]H(C)$. Second-period effort is increasing in C.

(iii) $U'(C) = -1 + [\alpha/(1+\alpha)][4K/\lambda(1-\lambda)]H(C) < 0$, $U''(C) = [\alpha/(1+\alpha)][4K/\lambda(1-\lambda)]H'(C) > 0$, and $U(C) > 0$ for $C < C_{max}$ and $U(C_{max}) = 0$.

(iv) $[d(\eta_i^f(C) - \eta_i^*(C))/d\lambda] < 0$.

(v) $[d(\eta_i^f(C) - \eta_i^*(C))/d\alpha] > 0$.

(vi) $\gamma(C) = 0\ \forall C < C_{max}$.

Proof: See the appendix.

The restriction above requiring $\psi(\bullet)$ to be quadratic is not necessary but allows a sufficiency theorem from Takayama (1985) to be applied. In the one-period model of Laffont and Tirole (1986), the condition $\psi''' > 0$ is sufficient. In this model, because of the plus term in the local incentive compatibility condition, a stricter condition is required. The form of the stricter condition will not be pursued here; instead, it will be assumed that $\psi(\bullet)$ is quadratic.

Using Proposition 2, part (vi), Eq. (8) states that first-period effort will equal the first-best level minus some amount that increases in C. Thus, first-period effort decreases in reported cost. In Eq. (9), second-period effort is equal to the first-best level plus some amount that increases in C. Thus, the government requires greater than first-best effort from the contractor in the second period—to punish severely the firm for any misreporting that it may do initially. Again, a low-cost firm that submits a high-cost report has a higher rate of cost at time λ than its reported type $\hat{C}$, and thus must "catch up" in the second period. If super-normal effort is already required by the government, then catching up is extremely costly to the firm. Thus, requiring super-normal effort in the second period cuts rents by making misreporting less attractive.

The intuition for parts (i) and (ii) in Proposition 2 is as follows. Profits for the firm are obtained by integrating the (ICF) constraint, or $U(C) = \int_C^{C_{\max}} [\psi'(\eta_1(z)) - \psi'(\eta_2(z))] \, dz + U(C_{\max})$. First-best effort has $\eta_1^f > \eta_2^f$ (see Proposition 1), and thus firms would earn rents if the first-best scheme is used. Note that a big difference between η_1 and η_2 by a high-cost firm is reflected in a higher profit for all firms with lower costs, so effort near the first-best level should be encouraged more for low-cost firms. For a given type C, gains from effort to DoD are the savings on procurement cost $(1 + \alpha)f(C)[(\lambda^2/2)\eta_1 + \lambda(1-\lambda)\eta_1 - \lambda\psi(\eta_1) + [(1-\lambda)^2/2]\eta_2 - (1-\lambda)\psi(\eta_2)]$, while the losses for a type C are profits to all of the firms "below" C, or $\alpha F(C)[\psi'(\eta_1) - \psi'(\eta_2)]$. Since $F(C)/f(C)$ is strictly increasing, this implies that $\eta_1^*(C) - \eta_2^*(C)$ shrinks as C increases.

Parts (i) and (ii) plus (IR) imply part (iii). Part (iv) implies that a shorter time period to the first audit results in a more drastic distortion of η_1^* from the first-best level. The cost of distorting away from first-best effort is less (because the cumulative effects of effort distortions are less over a short period), and so a larger distortion is induced, which makes concealment more expensive for the firm. Part (v) states that the size of the distortion in effort is increasing in the cost of transfers α. For example, for small α, the government does not mind large profits because, by assumption, it is relatively cheap to raise the funds to pay the profits. In that case, large distortions in effort away from the first-best that create a relatively large deadweight loss will be avoided.

Proposition 3 states the optimal transfer function for the case of $U' < 0$ (or "unconstrained" case), a single firm, and no uncertainty in costs.

Proposition 3: Define

$$S(\hat{C}) \equiv \lambda\psi(\eta_1^*(\hat{C})) + (1 - \lambda)\psi(\eta_2^*(\hat{C}))$$

$$+ \int_{\hat{C}}^{C_{\max}} [2\psi'(\eta_1^*(C)) - 2\psi'(\eta_2^*(C))] \, dC.$$

Under the assumptions of Proposition 2, the transfer function $R(\hat{C}, C_1, C_2) = S(\hat{C}) + C_1 + C_2$ (if and only if $C_1 = \hat{C}_1$ and $C_2 = \hat{C}_2$) implements the solution in Proposition 2.

Proof: See the appendix.

2.A.4.2. The constrained case. In the unconstrained case, $(d\eta_1^*/dC) < 0$ and $(d\eta_2^*/dC) > 0$ and $\eta_1^*(C_{min}) > \eta_2^*(C_{min})$. For C sufficiently large, $\eta_1^*(C)$ and $\eta_2^*(C)$ will be equal and thus, from (ICF), $U'(C) = 0$. Let $C_z \equiv \inf\{C \mid U'(C) = 0\}$. From Proposition 2, it is easily seen that $H(C_z) = [(1 + \alpha)/\alpha][\lambda(1 - \lambda)/4K]$. For $C < C_z$, the behavior of the solution is exactly as in Proposition 2. For $C \geq C_z$, a theorem from Baron and Besanko (1987b) (Appendix A) may be used to show that for all C, $U''(C) \geq 0$. This weak-convexity condition is sufficient to imply that $U(C) = U'(C) = 0$ for all $C \geq C_z$ (see Figure 8.3 for the diagram and the following proposition for the proof). These results are summarized in Proposition 4 and used in Section 3 to compare the efficiency of various contracts.

Proposition 4: Assume that $\psi(\bullet)$ is quadratic and that $H(C_{max}) \geq [(1 + \alpha)/\alpha][\lambda(1 - \lambda)/4K]$. Then the constrained case holds and the optimal solution is characterized by:

(i) For $C < C_z$, all of the results from Proposition 2 hold.

(ii) For $C \geq C_z$, $\eta_1^*(C) = \eta_2^*(C) = (1/2K)$. Also, $U(C) = 0$ and $U'(C) = U''(C) = 0$.

(iii) The transfer function $R(\hat{C}, C_1, C_2) = S(\hat{C}) + C_1 + C_2$ (if and only if $C_1 = \hat{C}_1$ and $C_2 = \hat{C}_2$), with $S(\hat{C})$ defined in Proposition 3, implements the solution.

Proof: See the appendix.

Total savings from cost reduction for firms with $C > C_{min}$ are strictly less than first best. Thus, the optimal solution trades some loss in cost reduction for some reductions in the rents to the firm.
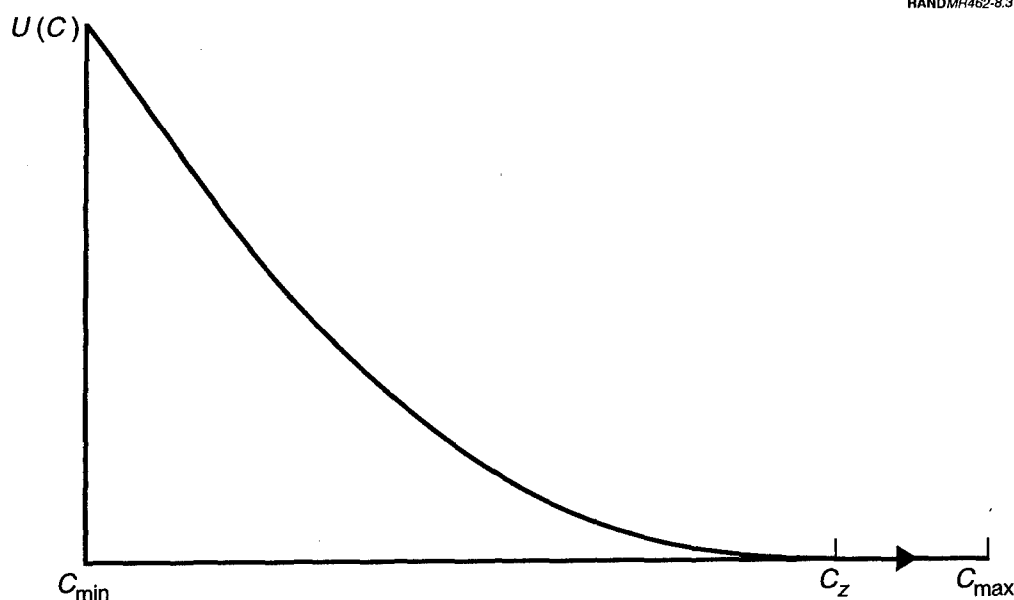
Figure 8.3—The Constrained Solution

2.A.5. Implementing the optimal contract in the context of DoD profit policy.

A profit policy contract specifies a lump-sum transfer plus markups $m_1(\hat{C})$ and $m_2(\hat{C})$ that apply to cost overruns and underruns: $R(\hat{C}, C_1, C_2) = S(\hat{C}) + C_1 + C_2 + m_1(\hat{C})[C_1 - \hat{C}_1] + m_2(\hat{C})[C_2 - \hat{C}_2]$, where the terms in brackets are the cost overruns or underruns.[16] The markups induce effort $\eta_1^*(\hat{C})$ and $\eta_2^*(\hat{C})$. The markups $m_1(\hat{C})$ and $m_2(\hat{C})$ will now be derived, assuming a uniform distribution over $[C_{min}, C_{max}]$, a quadratic cost function $\psi(\eta_i) = (K/2)\eta_i^2$, and $\alpha = \infty$ (this corresponds to the case in which DoD minimizes costs). Equations (8) and (9) then specify optimal effort as

$$\eta_1^*(C) = \frac{2-\lambda}{2K} - \frac{2}{\lambda}(C - C_{min})$$

$$\eta_2^*(C) = \frac{1-\lambda}{2K} - \frac{2}{1-\lambda}(C - C_{min})$$

if $C \le C_z = C_{min} + [\lambda(1-\lambda)/4K]$. The benefits of exerting $\eta_2(C)$ are $[(1-\lambda)^2/2]\eta_2 - (1+m_2)[(1-\lambda)^2/2]\eta_2$, and the costs are $(1-\lambda)(K/2)\eta_2^2$. The firm chooses η_2 to maximize net benefits. Equating marginal benefit to marginal cost, and substituting for $\eta_2(C)$, yields the markup that induces the proper level of effort $\eta_2(\hat{C})$:

$$m_2^*(\hat{C}) = -\frac{4K}{(1-\lambda)^2}(\hat{C} - C_{min}) - 1. \tag{10}$$

If $C \ge C_z$ then $m_2^*(\hat{C}) = -[4K/(1-\lambda)^2](\lambda(1-\lambda)/4K) - 1 = -[1/(1-\lambda)]$. For first-period effort, the benefits are $(\lambda^2/2)\eta_1 + \lambda(1-\lambda)\eta_1 - (1+m_1)(\lambda^2/2)\eta_1 - (1+m_2)(1-\lambda)\eta_1$, and costs are $\lambda(K/2)\eta_1^2$. The first-order condition yields

$$m_1^*(\hat{C}) = \frac{4K(1+\lambda)}{\lambda^2(1-\lambda)}(\hat{C} - C_{min}) - 1. \tag{11}$$

If $C \ge C_z$ then, substituting C_z in for C, $m_1^*(\hat{C}) = 1/\lambda$. Observe that m_1^* is increasing in $\hat{C}$ and that, for $\hat{C} > [\lambda^2(1-\lambda)/4K(1-\lambda)] + C_{min}$, $m_1^* > 0$ and the firm is thus paid extra *profit* for overruns. Of course, this profit is taken away at the end of the second period because first-period cost overruns carry over to the second period, where they are penalized heavily by a very low m_2^*.[17]

Equation (10) shows that the second-period markup is negative, i.e., the government not only does not reimburse the firm, but *charges* it a percentage of its cost overruns. These extreme incentives induce effort in excess of the first-best level in the second period. The result in Eq. (11) is roughly similar to a result in Bower and Osband. For a discussion of its comparative statics, see Chapter Four.

---

[16]The analysis here also could determine "cost-share" parameters, rather than markups. For those who prefer cost shares, substitute $m = -\theta$, where θ is the cost share.

[17]Markups in DoD have much less variability than identified as optimal by this model, possibly because political considerations and notions of "fairness" prevent highly uneven payments over time.

2.A.6. Noise costs. The analysis so far has been, using the terminology of Rogerson (1987), for a "standard" self-selection problem. That is, the model has no randomness in costs or in the audit. In this subsection "generalized" cost functions of the form $\tilde{C}_1 = C_1 + \tilde{\varepsilon}_1$, and $\tilde{C}_2 = C_2 + \tilde{\varepsilon}_2$, in which $\tilde{\varepsilon}_i$ are random variables with mean zero, are assumed. The government observes the realization $C_1 + \tilde{\varepsilon}_1$ of $\tilde{C}_1$ at $t = \lambda$, and the realization $C_2 + \tilde{\varepsilon}_2$ of $\tilde{C}_2$ at $t = 1$, and bases its reimbursements upon $\tilde{R}(\hat{C}, C_1 + \tilde{\varepsilon}_1, C_2 + \tilde{\varepsilon}_2)$.

In 2.A.4 and 2.A.5, the government implemented the contract: pay $R(\hat{C}, C_1, C_2)$ if audited costs $\tilde{C}_1 = \hat{C}_1$ and $\tilde{C}_2 = \hat{C}_2$, and $-\infty$ otherwise. Because of the randomness in costs (or noise in the audit), this contract is no longer feasible, as the firm would be unwilling to participate. Given the introduction of noise, the optimal form of $\tilde{R}(\bullet)$ is not transparent. However, it can be shown that a particularly simple form of $\tilde{R}(\bullet)$ is optimal: The transfer should be $\tilde{R}(\hat{C}, C_1 + \tilde{\varepsilon}_1, C_2 + \tilde{\varepsilon}_2) = S(\hat{C}) + C_1 + \tilde{\varepsilon}_1 + C_2 + \tilde{\varepsilon}_2 + m_1(\hat{C})(C_1 + \tilde{\varepsilon}_1 - \hat{C}_1) + m_2(\hat{C})(C_2 + \tilde{\varepsilon}_2 - \hat{C}_2)$, in which the lump-sums $S(\hat{C})$ and the markups m_1 and m_2 are derived from the solution to the standard problem identified above. Thus, a key result of Laffont and Tirole (1986) generalizes to this model.

Proposition 5: The solution (a menu of linear contracts) identified for the standard self-selection problem is optimal for the generalized problem.

Proof: See the appendix.

2.A.7. Multiple firms. The preceding analysis assumes a single firm; now assume N identical firms in the industry. Laffont and Tirole (1987) have shown that it is easy to transfer the result to the case of multiple firms with independent and identically distributed costs. In this case, the winner receives a profit of $-\int_{\hat{C}_1}^{\hat{C}^2} U'(C)dC$, instead of $-\int_{\hat{C}^1}^{C^{max}} U'(C)dC$, where $\hat{C}^i$ is the ith lowest report. From the winning firm's perspective, $\hat{C}^2$ is exogenous, so incentive compatibility is maintained, regardless of the strategies of the other firms.

2.B. Commitment Plus Renegotiation

2.B.1. Basic model. The full-commitment solution requires greater than first-best effort in the second period, so an opportunity for *ex post* renegotiation of the contract arises.[18] In particular, at time λ a Pareto-improving contract can be written in which effort is set at the first-best level. A lack of credible commitment not to renegotiate at time λ will preclude implementation of the original contract. In this "limited commitment" case, the types of renegotiation must be clarified. Here, it will be assumed that the government may credibly commit to a particular transfer function $R(\hat{C}^1, \hat{C}^2, C_1, C_2) = S(\hat{C}^1, \hat{C}^2) + C_1 + C_2 + m_1(\hat{C}^1)(C_1 - \hat{C}_1) + m_2(\hat{C}_2)(C_2 - \hat{C}_2)$. This is the status quo contract. The firm may also commit to that contract. However, both parties may agree at time λ to voluntarily renegotiate.

---

[18]In equilibrium, the firm's report at $t = 0$ has revealed its cost. It is assumed that the government may at least commit to its mechanism for the period $t \in [0, \lambda]$.

Suppose the parties cannot commit not to renegotiate the original contract at time λ. This modifies the objective function in P1 and also changes the (IC) constraint, as the firm knows that the contract is renegotiable and takes that into account when it reports its cost. In this subsection, a separating contract will be constructed that entices the firm to tell the truth about its cost at time 0 and that incorporates a fixed-price contract after time λ. Laffont and Tirole (1988b) show in a similar model that a separating contract is never optimal, and Baron and Besanko (1987a) identify a pooling equilibrium that marginally improves welfare. The idea is that a pooling equilibrium loses some efficiency because the policy is not completely responsive to type but saves *ex ante* on information rents that must be given to the firm to reveal its type. However, as in the example in Baron and Besanko, the examples in Section 3 suggest that the benefits to the full-commitment contract relative to even this fully separating, limited-commitment contract will be very small in percentage terms. Nevertheless, the contract derived in 2.B.3. is not the optimal contract. Hence, the solution for the limited commitment case that will be presented here, namely, a separating contract, identifies only an upper bound on the losses identified in Section 3 because of the limited commitment powers of the government. The derivation of the solution is obtained using the same methods as in subsection 2.A. The solution to the monopoly, no-noise case easily generalizes to the multiple-firm, noisy-monitor case.

2.B.2. Incentive compatibility condition. The derivation of the incentive compatibility condition proceeds along the same lines as in Section 2.A, with the additional constraint that $\eta_2(C) = \eta_2^f$. This yields a condition of

$$U'(C) = -2\psi'(\eta_1(C)) + 1 - \lambda. \tag{ICL}$$

Note that $2\psi'(\eta_2(C)) = 1 - \lambda$ if η_2 is set at the first-best level, so (ICL) is of the same form as (ICF). The term $1 - \lambda$ arises because of the effect outlined in Figure 8.2. However, the rents are not decreased by as much as in the full commitment case, because the government can no longer credibly commit to extracting super-normal effort in the second period. This makes it easier for the firm to conceal misreporting in the first period with hard work in the second.

2.B.3. Solution with commitment and renegotiation. This section solves for the optimal fully separating contract under commitment and renegotiation. The government's problem is

$$\min \int_{C_{\min}}^{C_{\max}} \left\{ (1+\alpha) \left[\lambda \psi(\eta_1(C)) - \frac{\lambda^2}{2} \eta_1(C) - \lambda(1-\lambda)\eta_1(C) \right. \right.$$

$$\left. \left. + (1-\lambda)\psi\left(\eta_2^f\right) - \frac{(1-\lambda)^2}{2}\eta_2^f + C \right] + \alpha U(C) \right\} f(C)\, dC \tag{P2}$$

$$\text{s.t.} \quad U'(C) = -2\psi'(\eta_1(C)) + 1 - \lambda \tag{ICL}$$

$$U(C) \geq U(\hat{C} \mid C), \quad \forall \hat{C} \tag{GICL}$$

$$U(C) \geq 0. \tag{IR}$$

The limited-commitment solution is similar to the results for the full-commitment solution. Proposition 6 is the analog of Proposition 4 for the limited-commitment case. Define C_y as inf $\{C \mid U'(C) = 0\}$.

Proposition 6: Assume a quadratic ψ and limited commitment. Then

(i) For $C < C_y$, $\eta_1^*(C) = [(2 - \lambda)/2K] - [\alpha/(1 + \alpha)](2/\lambda)H(C)$ and $\eta_2^*(C) = [(1 - \lambda)/2K]$. Profit $U(C) > 0$ and $U'(C) = -1 + [\alpha/(1 + \alpha)](4K/\lambda)H(C) < 0$ and $U''(C) = [\alpha/(1 + \alpha)](4K/\lambda)H'(C) > 0$.

(ii) For $C \geq C_y$, $\eta_1^*(C) = \eta_2^*(C) = [(1 - \lambda)/2K]$. Profit $U(C) = 0$ and $U'(C) = U''(C) = 0$.

(iii) Profits are equal or higher for the firm under renegotiation than under full commitment.

Proof: See the appendix.

The markups associated with the limited-commitment solution can be easily derived, but are omitted here.

3. COMPARISONS

3.A. Performance Comparisons

The solutions to two procurement models have been derived here. Bower and Osband provide a solution to a third model, which involves a single (optimal) incentive contract (i.e., the government offers an incentive contract with the same markup for every type of firm) in the first period and a fixed-price contract in the second period. It has been shown that the full-commitment menu does best and the limited-commitment menu does second best. The Bower-Osband contract is just the limited-commitment contract with the additional restriction of no menus, and so it does third best. However, besides the ordinal ranking of these contracts, the magnitudes of the gains from using these more complicated schemes are of interest. It is difficult to obtain analytic solutions, as can be verified from inspecting Eqs. (8) and (9) and Proposition 3. Therefore, numerical solutions to examples have been used to compare contract efficiency.[19] I will compare these three schemes to two simpler schemes—a pure fixed-price contract and a pure cost-plus contract. The three schemes previously characterized can be viewed as increasingly sophisticated combinations of fixed-price and cost-plus contracts. Thus, we have five contracts to compare, plus a sixth type of contract, the full-information, first-best contract in which the government pays the low-cost firm its costs, and extracts first-best effort as well.

---

[19]Several results have been verified analytically for the uniform-quadratic case.

For illustrative purposes consider two examples. Assume a uniform density, $\alpha = \infty$ (thus the government wishes to minimize total project costs), a quadratic effort function with constant $K = .01$, and $\lambda = .5$. First-best effort yields total cost reduction of 31.25. Table 8.1 shows expected total cost to the government. All expectations in this section are taken with respect to $F(\bullet)$; i.e., the government's expected cost, before it learns the reports of the firms. In the case of costs distributed uniformly in the interval [100, 130] and $N = 2$, one can see that the menu of contracts with commitment (FC) with an expected cost of 97.27, the menu with limited commitment (LC) (98.51), and the Bower-Osband solution (SI)(100.31) differ by little, although all three do noticeably better than the cost-plus-fixed-fee (CPFF) (110) (the fixed fee here is zero for comparative purposes) or the fixed-price (FP) (104.38) contract. The fixed-price contract does better than the cost-plus because the possible range of costs [100, 130] is relatively narrow, thus losses to adverse selection are small. Both the FC and LC solutions generate slightly more cost reductions than the SI solution. Again, the LC cost is an upper bound on the costs to the government under renegotiation.

If the range of costs is wider ([100, 200]), then the cost-plus contract outperforms the fixed-price. Again, in this case, all three of the more sophisticated schemes outperform the simple contracts, but the difference between the single-incentive contract (123.69) and FC scheme (120.69) is only 2.6 percent. Since commitment not to renegotiate is unlikely, particularly in a long-term DoD engagement, the better comparison may be SI and LC (123.45). They differ by only 0.4 percent! In the example with costs in [100, 130], the gain is 3.1 percent to FC and 1.9 percent to LC when compared to the single-incentive contract. The gain to using the more sophisticated menu of

Table 8.1

Uniform Distribution, $K = .01$, $\lambda = 0.5$, $\alpha = \infty$

Contract	[100, 130]				[100, 200]			
	Prod. Cost	Firm Profit	Total Cost	%[a]	Prod. Cost	Firm Profit	Total Cost	%[a]
	N = 1							
FB	99.38	0	99.38	50	134.38	0	134.38	50
FC	102.07	.10	102.17	.41	137.31	.03	137.34	47
LC	103.89	.87	104.76	32	140.11	.26	140.37	44
SI	105.63	0	105.63	29	140.63	0	140.63	44
CPFF	115	0	115	-2	150	0	150	34
EP	99.38	15	114.38	0	134.38	50	184.38	0
	N = 2							
FB	94.38	0	94.38	67	117.71	0	117.71	67
FC	96.77	.5	97.27	57	120.42	.27	120.69	64
LC	97.01	1.5	98.51	53	122.94	.51	123.45	61
SI	98.31	2.0	100.31	47	123.69	0	123.69	61
CPFF	110	0	110	15	133.33	0	133.33	51
EP	94.38	10	104.38	33	117.71	33.33	151.04	33

[a]Percentage of potential savings, with FP, $N = 1$ total costs defined as 0 percent and FP, $N = \infty$ total costs defined as 100 percent.

contracts is that they can "fine-tune" the tradeoff between moral hazard and adverse selection. Since adverse selection dominates moral hazard in the example with costs in [100, 200], schemes that trade off the two tend to perform about the same.[20]

The closeness in expected total procurement cost does not change for these examples even if there is only one firm. Observe that, in the example with costs in [100, 130], LC and SI actually become closer together as N decreases. The intuition is in Section 3.B.1.

Another way to measure performance of the contracts is to compare them to two yardstick contracts: first, the fixed-price contract with one firm in the industry and second, the fixed-price contract with an infinite number of firms (which yields C_{min} less first-best efficiency gains). Cost savings relative to the monopoly contract are measured as a percentage of cost savings achieved by the perfect-competition contract and shown in Table 8.1. For example, with two firms, a fixed-price contract garners one-third of the gains, since $E(C^2)$ lies one-third of the way from the monopoly contract toward the perfect competition contract. If costs are distributed in [100, 200], then a cost-plus contract manages to extract 51 percent of the gains, and FC, LC, and SI capture over 60 percent of the gains. In the example with costs in [100, 130], there is more of a tradeoff between the two informational asymmetries, and as a result the more finely crafted FC and LC schemes perform relatively better than the single-incentive contract.

3.B. Value of Procurement Instruments

All of the procurement instruments available cost the government something to employ. This subsection identifies the benefits of each instrument. Those benefits can then be weighed against the potential costs to serve as guidelines in determining policy.

The benefits are identified by measuring performance differences among different contracts. Of course, the size of the benefits depends on the values of the parameters. For example, the value of a menu depends upon the value of seven other parameters: C_{min}, C_{max}, N, α, λ, the ability or inability to commit, and the number of audits conducted. Nevertheless, comparisons of examples reveal, to a reasonable approximation, the value of each procurement instrument as a function of the parameters. Space considerations prevent the enumeration of many of the parameter combinations, but some of the most important and interesting combinations are shown in what follows.

3.B.1. Value of a menu. In all of these cases, the quantity procured is fixed at one unit, so the benefits from adjusting the quantity to the firm's costs are not present. The value of the single-incentive contract scheme thus comes from using one incentive contract that trades off adverse selection and moral hazard. The limited-commitment schemes use a menu of incentive contracts from which the firm picks.

---

[20]The examples above assume $\alpha = \infty$—perhaps the differences between contracts are larger for lower α. However, as $\alpha \to 0$, all of the incentive contracts, in the limit, converge to a fixed-price contract.

For this model, this self-selection feature constitutes a second-order refinement of the contracts, and consequently cost savings are typically small. In the examples, a single incentive contract performs quite a bit better than a fixed-price or cost-plus contract, and 98 to 99.8 percent as well as a menu of contracts. Apparently, at least in the uniform-quadratic case, one incentive contract does almost as well as an infinite number. Furthermore, there have been no examples discovered to date in which the value of a menu exceeded 2 percent of procurement cost.

The value of a menu has an interior maximum with respect to the size of the support of the distribution, holding other parameters constant. See Figure 8.4.[21] If $C_{max} - C_{min}$ is zero, then a fixed-price contract is optimal and the value of a menu is zero. If $C_{max} - C_{min}$ is very large, then (almost) all firms are in the range in which $U(C) = 0$ and the SI contract can implement that solution, and a menu is not necessary. The value of a menu attains its maximum for some intermediate value of $C_{max} - C_{min}$. One implication of this result is that the value of a menu may be *higher* for multiple firms than for just one firm. That is, perhaps somewhat counterintuitively, the gains to this contract refinement may not be largest for the monopoly case. The reason for the result can be deduced from inspection of the optimal solution for the uniform-quadratic case. See Figure 8.5. Observe that, for all $C \geq C_z$, a single incentive contract is used; it is only in the region $C < C_z$ that a menu is actually used. Therefore, if C_{max} is much larger than C_z, and there is only one firm, then a menu is worth little.
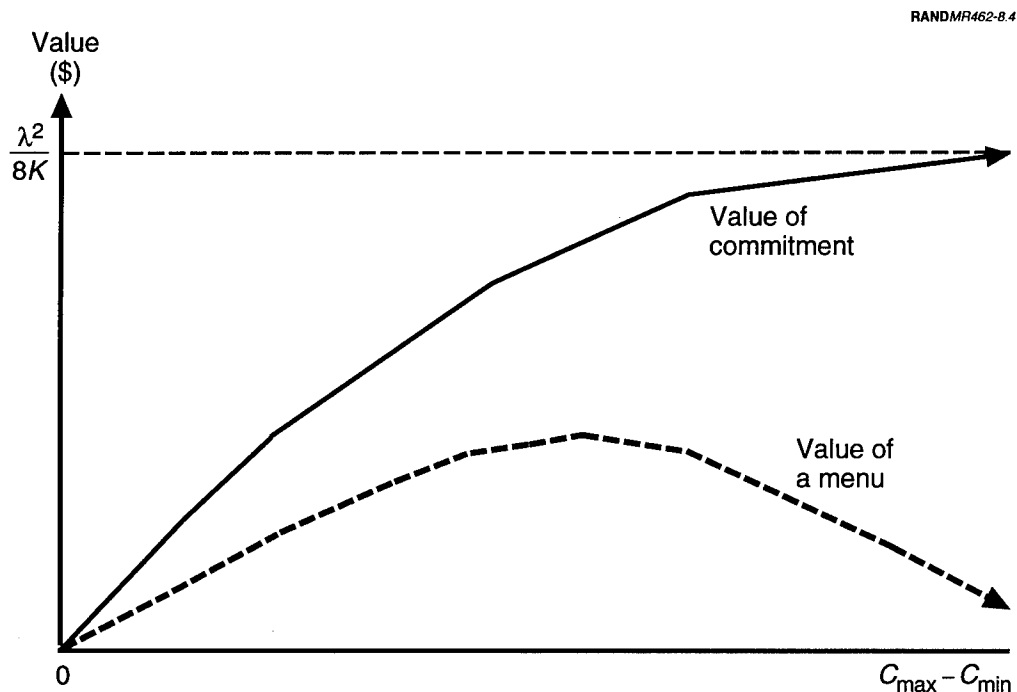
RANDMR462-8.4

Figure 8.4—Value of a Menu and Commitment (uniform $f(\bullet)$)

---

[21] The result has been verified analytically for the uniform-quadratic case.
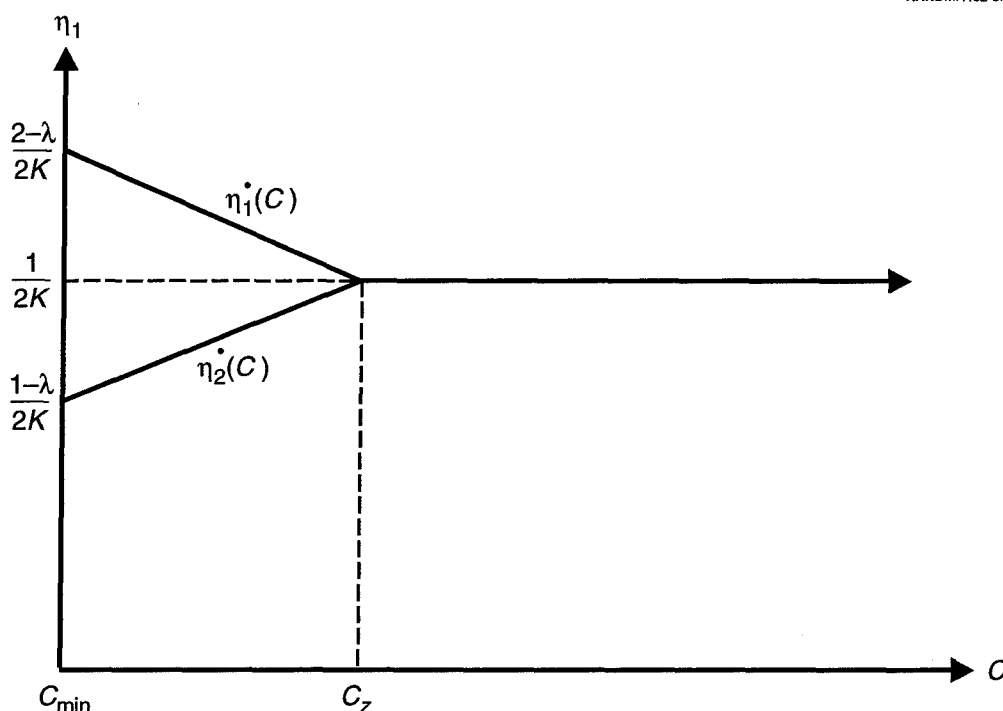
Figure 8.5—Solution to the Full-Commitment, Uniform-Quadratic Case

As N increases, the distribution of the first-order statistic C^1 becomes more concentrated at the lower end of the cost distribution and a menu is thus used more. However, as $N \to \infty$, the distribution of C^1 converges to a point mass at C_{min}, where a fixed-price contract can be used, and so the value of a menu decreases to zero as $N \to \infty$.

3. B. 2. Value of full commitment. Comparing FC to LC in Table 8.1, note that the loss from an inability to commit at time $\lambda = 0.5$ is (at most) quite small (since LC is an upper bound). Thus, in the examples, the inability to commit to a long-term policy does *not* hurt the government very much. The results suggest that a fixed-price contract after the audit is completed does quite well, and, of course, is easier to administer and invulnerable to renegotiation because effort in the second period is first best.[22]

Even the limited-commitment solution assumes that the government commits to a first-period markup. If the government cannot commit to even a first-period markup, then only a first-price auction can be used. Comparing FP to LC, we see that the ability to commit at least initially is of substantial value. Thus, loosely, the table identifies diminishing marginal returns to commitment.

---

[22]Furthermore, a referee has pointed out that long-term commitments may be unrealistic or undesirable because of a rapidly changing political or technological environment.

Returning to Figure 8.4, it can be shown in the uniform-quadratic case that the value of full commitment rises to a maximum of $\lambda^2 / 8K$. The value of full commitment is derived from the ability to distort high-types' effort in the second period up to $1/2K$. As $C_{max} - C_{min}$ increases, all types are given the incentive to do this (see Figure 8.5) and the value asymptotes to $\lambda^2 / 8K$.

3.B.3. Value of auditing.

Similarly, the examples also tell us something about the value of auditing. The contract FC requires auditing at $t = \lambda$ and $t = 1$. The contract LC may be interpreted as a contract at which auditing at $t = 1$ is not possible, and FP as a contract in which no auditing at all is possible. Thus, the gains to auditing once are fairly large (5.6 through 18.3 percent in the examples), but the marginal gains to auditing twice shrink drastically (an additional 1.2 to 1.8 percent). This diminishing-marginal-returns effect is not as strong if the distribution of costs is narrow; then gains to both audits are quite small. It is conjectured with some confidence that in this model the marginal gains from additional audits beyond two are negligible. It would appear in practice that the costs of additional monitoring and bureaucracy would quickly outstrip the benefits from keeping close watch over the firm. Thus, given a fixed departmental budget for auditing, DoD should not allocate auditors to audit projects repeatedly unless major new informational asymmetries evolve over time. Thus the model confirms our intuition. Naturally, the assumption that the underlying cost C (which effort modifies) does not change plays a big role here; in the real world, design changes may alter the cost structure and introduce new informational asymmetries over time.[23] In this model, the only reason to audit is to ameliorate the original information asymmetry.

Auditing adds the most if matched with a menu of contracts and full commitment, because the contract fully utilizes the information gleaned from the audit. Finally, the value of auditing is increasing in the uncertainty of costs; the wider the possible range of costs, the more valuable the audit.

3.B.4. Value of competition.

The gains from carefully calibrating the incentive contract versus the gains from increasing competition are of interest. The examples suggest that competition is much more valuable. Figure 8.6 shows the percentage of potential savings for one to four identical firms. Here, $K = .005$ and $\lambda = 0.5$. In this example, because the support of the cost distribution ([108, 120]) is fairly narrow, the gains to using the most sophisticated schemes are largest for the monopoly case. As N increases, note that all of the schemes tend to converge in performance and then converge more slowly to the first best. At $N = 3$, the fixed-price contract does almost as well as any of the other schemes. It seems that competition tends to cure all inefficiencies in the contracting process. In particular, a fixed-price contract does almost as well as any other contract as N increases.[24] The marginal value of other procure-

---

[23]For example, General Dynamics, the prime contractor for the F-111 program, experienced an average of 750 contract changes per year from the Air Force on the project, many of those Engineering Change Proposals (ECPs) (Fox, 1974, p. 364).

[24]Furthermore, fixed-price contracts are also cheaper to administer. However, some case studies suggest that fixed-price contracts cannot be implemented in complex weapons systems procurements (Fox, 1974). The inability to use fixed-price contracts may have an ambiguous effect on the value of competition.
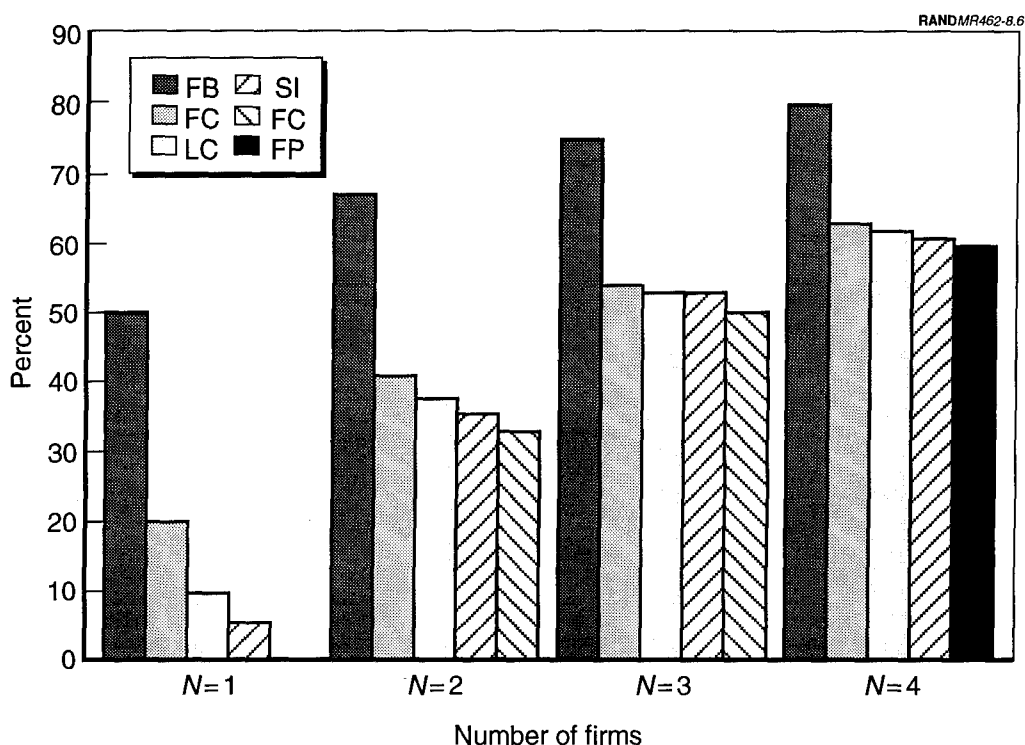
RANDMR462-8.6

Figure 8.6—Percentage of Potential Savings, 1–4 Firms
$(U \sim [108, 120], N = 1 - 4, K = .005)$

ment instruments eventually declines to zero as competition increases. Furthermore, in the examples, gains to competition are much larger than the gains to using a menu of contracts—for example, compare SI at $N = 1$ to LC at $N = 1$ and SI at $N = 2$—when utilizing a single incentive contract, the gain to adding a firm is much larger than the gain to switching to the more finely tuned limited-commitment contract. Competition, in the long run, may be stimulated by increasing R&D funding, or perhaps, by simply encouraging more firms to bid for a contract. Two major caveats apply, however. First, the analysis ignores bid preparation costs incurred by prospective contracts. Second, typically the marginal competitor does not have identical prospects to the other firms, but instead has a less-efficient technology. Both of these effects lessen the gains to competition.[25]

4. CONCLUSIONS AND EXTENSIONS

I have explored a two-stage procurement model with hidden action, hidden information, and discretionary learning. Adding the aspect of learning changes the optimal solution with commitment away from the repeated one-shot solution, as first-period actions now affect the second-period cost structure of the firm. With learning

---

[25]For another paper that finds substantial advantages to competition over self-selection, see Osband (1989b).

and total-cost auditing, charging the firm (i.e., having a markup less than -1) in the second period reduces the incentive to misreport cost in the first period.

A major aim of this paper was to quantify the magnitude of the differences among successively more fine-tuned contracts. In the examples, the gains from using an incentive contract instead of a fixed-price or cost-plus contract are fairly sizable—10 percent or so—but further gains from using a menu of incentive contracts are much smaller, typically only 0 to 2 percent. Furthermore, menus are more difficult to implement. The gains from full commitment were typically 1 to 3 percent of total cost, and multiple audits showed similar gains. The examples show that gains to increased competition outstrip the gains available from any other improvement.

Several extensions of the model are conceivable. For example, firms may have a better idea of rivals' costs than does the government (possibly because costs are correlated). Correlation changes the optimal contract, and it seems unclear whether this would decrease the value of competition (firms would be more alike, which reduces the value of an additional draw, but also allows the principal to economize on information rents, since a firm's report reveals information about other firms' costs). Another extension might be to allow the disutility of effort K to be the private information of the firm's manager, while assuming that costs are observed from a previous relationship. This framework is similar to the model studied here, and might yield a similar type of solution. A third extension is to compute numerical examples for cost distributions other than the uniform distribution.

APPENDIX

Proof of Proposition 2: First Theorem 8.C.5 (p. 662) in Takayama (1985) shows that the necessary conditions for an optimum are also sufficient if the objective function and the state equation (ICF) are convex in (η_1, η_2, U). In P1 the objective function is strictly convex in η_1 and η_2 and linear in U. The state equation (ICF) is convex if and only if $\psi''' = 0$ (hence the restriction to a quadratic $\psi(\bullet)$). Thus the necessary conditions (6a) through (6d) are sufficient for an optimum.

(i) Note that $\psi'(\eta_1^*(C_{min})) = [(2 - \lambda)/2]$ and $\psi'(\eta_2^*(C_{min})) = [(1 - \lambda)/2]$ and therefore $U'(C_{min}) < 0$ and $\gamma(C_{min}) = 0$. Furthermore, the assumption $U'(C) < 0$ implies $U(C) > 0 \ \forall C < C_{max}$. And from Eq. (6d) then $\gamma(C) = 0 \ \forall C < C_{max}$, which establishes part (vi). The equations for $\eta_i^*(C)$ are found from Eqs. (8) and (9). Define $\psi_i \equiv \psi(\eta_i(C))$. Differentiate Eq. (8) (using $\gamma(C) = 0$) with respect to C to obtain

$$\psi_1'' \frac{d\eta_1^*(C)}{dC} = -\frac{\alpha}{1+\alpha} \frac{2}{\lambda} \left[H'\psi_1'' + H\psi_1''' \frac{d\eta_1^*}{dC} \right].$$

Since $\psi''' = 0$ and $\psi_1'' = K$ this yields

$$\frac{d\eta_1^*(C)}{dC} = -\frac{\alpha}{1+\alpha} \frac{2}{\lambda} H' < 0.$$

(ii) Differentiate Eq. (9) to obtain

$$\psi_2'' \frac{d\eta_2^*(C)}{dC} = \frac{\alpha}{1+\alpha} \frac{2}{1-\lambda} H'\psi_2'' + \frac{\alpha}{1+\alpha} \frac{2}{1-\lambda} H\psi_2''' \frac{d\eta_2^*(C)}{dC},$$

which reduces to

$$\frac{d\eta_2^*(C)}{dC} = \frac{\alpha}{1+\alpha} \frac{2}{1-\lambda} H' > 0.$$

(iii) Differentiate (IFC) and use parts (i) and (ii) to obtain

$$U''(C) = -2\psi''(\eta_1^*(C)) \frac{\partial \eta_1^*(C)}{\partial C} + 2\psi''(\eta_2^*(C)) \frac{\partial \eta_2^*(C)}{\partial C} > 0,$$

so $U(C)$ is strictly convex. By assumption, $U'(C) < 0$ and thus $U(C_{max}) < U(C) \ \forall C < (C_{max})$. Individual rationality requires $U(C) \geq 0$, so it is optimal to set $U(C_{max}) = 0$. Thus $U(C) > 0$ for $C < C_{max}$.

(iv) and (v) Subtract $\psi'(\eta_1^*)$ from $\psi'(\eta_1^f)$ to obtain

$$\psi'(\eta_1^f(C)) - \psi'(\eta_1^*(C)) = \frac{\alpha}{1+\alpha} \frac{2}{\lambda} H(C)\psi_1''(\eta_1^*(C)). \tag{A.1}$$

Differentiate the right-hand side of Eq. (A.1) with respect to α to obtain part (iv) and λ to obtain part (v).

Proof of Proposition 3: Differentiate Eq. (4) evaluated at $\eta_1^*(C)$ and $\eta_2^*(C)$ with respect to $\hat{C}$, and use Eqs. (1) and (3) to get

$$\frac{\partial U(\hat{C} \mid C)}{\partial \hat{C}} = S'(\hat{C}) - \lambda\psi'\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)\left(\frac{\partial \eta_1^*(\hat{C})}{\partial \hat{C}} - \frac{2}{\lambda}\right)$$

$$- (1-\lambda)\psi'\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)\left(\frac{\partial \eta_2^*(\hat{C})}{\partial \hat{C}} + \frac{2}{1-\lambda}\right), \tag{A.2}$$

in which $S'(\hat{C}) = \lambda \psi'(\eta_1^*(\hat{C}))[\partial \eta_1^*(\hat{C}) / \partial \hat{C}] + (1 - \lambda)\psi'(\eta_2^*(\hat{C}))[\partial \eta_2^*(\hat{C}) / \partial \hat{C}] - 2\psi'(\eta_1^*(\hat{C})) + 2\psi'(\eta_2^*(\hat{C}))$. Note that $(\partial U / \partial \hat{C})|_{\hat{C}=C} = 0$, so the solution satisfies the first-order condition. For the local second-order condition, use the following identity:

$$\frac{\partial^2 U}{\partial \hat{C}^2}\bigg|_{\hat{C}=C} = -\frac{\partial^2 U}{\partial \hat{C}^2 \partial C}\bigg|_{\hat{C}=C}.$$

Differentiating Eq. (A.2) with respect to C and setting $\hat{C} = C$ yields

$$\frac{\partial^2 U}{\partial \hat{C} \partial C}\Big|_{\hat{C}=C} = 2\psi''(\eta_1^*(C))\left(\frac{\partial \eta_1^*(C)}{\partial C} - \frac{2}{\lambda}\right)$$
$$-2\psi''(\eta_2^*(C))\left(\frac{\partial \eta_2^*(C)}{\partial C} + \frac{2}{1-\lambda}\right). \tag{A.3}$$

From Proposition 2, parts (i) and (ii), it is established that $(\partial \eta_1^* / \partial C) < 0$, and $(\partial \eta_2^* / \partial C) > 0$, and thus the right-hand side of Eq. (A.3) is negative.

Global incentive compatibility requires that $U(C) \geq U(\hat{C}|C)$ for all $\hat{C}$, C. Substitute Eqs. (1) and (3) into Eq. (4) to obtain

$$U(\hat{C}|C) = S(\hat{C}) - \lambda\psi\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)$$
$$- (1-\lambda)\psi\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right), \tag{A.4}$$

where

$$S(\hat{C}) \equiv \lambda\psi(\eta_1^*(\hat{C})) + (1 - \lambda)\psi(\eta_2^*(\hat{C}))$$
$$+ \int_{\hat{C}}^{C_{max}} [2\psi'(\eta_1^*(C)) - 2\psi'(\eta_2^*(C))]\, dC. \tag{A.5}$$

Substitute Eq. (A.5) into Eq. (A.4) to obtain

$$U(\hat{C}|C) = \lambda\left(\psi(\eta_1^*(\hat{C})) - \psi\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)\right)$$
$$+ (1-\lambda)\left(\psi(\eta_2^*(\hat{C})) - \psi\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)\right)$$
$$+ \int_{\hat{C}}^{C_{max}} [2\psi'(\eta_1^*(C)) - 2\psi'(\eta_2^*(C))]\, dC. \tag{A.6}$$

Profits, given a truthful report C, are

$$U(C|C) \equiv U(C) = \int_{C}^{C_{max}} [2\psi'(\eta_1^*(C)) - 2\psi'(\eta_2^*(C))]\, dC.$$

Now, assume that $\hat{C} > C$ (the proof for $\hat{C} < C$ is similar and omitted). Then $U(C) \geq U(\hat{C}|C)$ if and only if

$$\int_C^{\hat{C}} [2\psi'(\eta_1(C)) - 2\psi'(\eta_2^*(C))] \, dC \geq \lambda\left(\psi(\eta_1^*(\hat{C})) - \psi\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)\right)$$

$$+ (1 - \lambda)\left(\psi(\eta_2^*(\hat{C})) - \psi\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)\right).$$

Observe that the left-hand side = right-hand side = 0 if $\hat{C} = C$. Then, it must be that if $(d\text{LHS}/d\hat{C}) \geq (d\text{RHS}/d\hat{C})$ for all $\hat{C} > C$, the inequality holds. Taking the derivative of both sides yields

$$2\psi'(\eta_1^*(\hat{C})) - 2\psi'(\eta_2^*(\hat{C}))$$

$$\geq 2\psi'\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right) - 2\psi'\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)$$

$$+ \lambda \frac{d\eta_1^*(\hat{C})}{d\hat{C}}\left(\psi'(\eta_1^*(\hat{C})) - \psi'\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)\right)$$

$$+ (1 - \lambda)\frac{d\eta_2^*(\hat{C})}{d\hat{C}}\left(\psi'(\eta_2^*(\hat{C})) - \psi'\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)\right).$$

Rearrange to obtain

$$\left(\psi'(\eta_1^*(\hat{C})) - \psi'\left(\eta_1^*(\hat{C}) + \frac{2}{\lambda}(C - \hat{C})\right)\right)\left(2 - \lambda\frac{d\eta_1^*(\hat{C})}{d\hat{C}}\right)$$

$$- \left(\psi'(\eta_2^*(\hat{C})) - \psi'\left(\eta_2^*(\hat{C}) - \frac{2}{1-\lambda}(C - \hat{C})\right)\right)\left(2 + (1 - \lambda)\frac{d\eta_2^*(\hat{C})}{d\hat{C}}\right) \geq 0. \qquad (A.7)$$

By assumption, ψ is convex and $\hat{C} > C$, thus $\psi'(\eta_1^*(\hat{C})) > \psi'(\eta_1^*(\hat{C}) + (2/\lambda)(C - \hat{C}))$ and $\psi'(\eta_2^*(\hat{C})) < \psi'\eta_2^*(\hat{C}) - [2/(1-\lambda)](C - \hat{C})$. Furthermore, it has been shown that $[d\eta_1^*(\hat{C})/d\hat{C}] < 0$ and $[d\eta_2^*(\hat{C})/d\hat{C}] > 0$. Thus, it can be seen that the inequality (A.7) holds. This completes the proof of global incentive compatibility.

Proof of Proposition 4:

(i) The proof is the same as in Proposition 2.

(ii) and (iii) Note that $\eta_1^*(C_z) = \eta_1^*(C_z) = 1/2K$. (It is easy to show that if $\eta_1 = \eta_2$ then $\eta_i = 1/2K$ is the most efficient choice of effort; that is, it maximizes net cost reductions.)

It is proposed that $U(C) = 0$ for $C \geq C_z$. Suppose not. Then, using the continuity of U, there exists $C_a > C_z$ where $U(C_a) > 0$ and $U'(C_a) > 0$. This implies that $\eta_2(C_a) > \eta_1(C_a)$. Using the same method of proof as in Appendix A of Baron and Besanko

(1987b), it can be shown that, for all C, $U''(C) \geq 0$.[26] Thus, $U(C)$ is increasing and weakly convex. So $U(C) > 0$, $\eta_1(C) < 1/2K$, and $\eta_2(C) > 1/2K$ for all $C \geq C_a$ and for any $C \in [C_z, C_a]$ such that $U(C) > 0$. Consider the following alternative scheme: Set $\eta_1(C) = \eta_2(C) = 1/2K$ and $U(C) = 0$ for all $C \geq C_z$. This scheme produces higher net cost reductions (because effort is distorted away from first best less under this scheme) and lower profits for each type $C \geq C_a$, and for any types $C \in [C_z, C_a]$ for which $U(C) > 0$. Thus, this scheme performs better than the scheme in which $U(C_a > 0$. Furthermore, it can be implemented using the transfer function stated in part (iv); $q(\hat{C}, C_1, C_2) = S(\hat{C}) + C_1 + C_2$. First, this scheme is incentive compatible: Check Eq. (A.2) using $[\partial \eta_i(\hat{C})/\partial C] = 0$, $S'(\hat{C}) = 0$ and $\eta_1^*(C) = \eta_2^*(C)$. Second, the scheme (weakly) satisfies the second-order conditions (A.3). Third, global incentive compatibility (A.7) is also satisfied.

This completes the proof.

Proof of Proposition 5: Consider a transfer function for the generalized problem of

$$\tilde{q}(\hat{C}, \tilde{C}_1, \tilde{C}_2) = S(\hat{C}) + C_1 + \tilde{\varepsilon}_1 + C_2 + \tilde{\varepsilon}_2 + m_1(\hat{C})(C_1 + \tilde{\varepsilon}_1 - \tilde{C}_1) + m_2(\hat{C})(C_2 + \tilde{\varepsilon}_2 - \tilde{C}_2),$$

which yields profits to the firm of

$$U(\hat{C}, \tilde{C}_1, \tilde{C}_2) = S(\hat{C}) + m_1(\hat{C})(C_1 + \tilde{\varepsilon}_1 - \hat{C}_1)$$
$$+ m_2(\hat{C})(C_2 + \tilde{\varepsilon}_2 - \hat{C}_2) - \lambda \psi(\eta_1) - (1 - \lambda)\psi(\eta_2). \tag{A.8}$$

The optimal solution in the standard problem requires

$$m_1^*(\hat{C}) = \frac{4\psi'(\eta_2^*(\hat{C})) - 2\psi'(\eta_1^*(\hat{C}))}{\lambda}$$

$$m_2^*(\hat{C}) = \frac{2\psi'(\eta_2^*(\hat{C}))}{1 - \lambda}. \tag{A.9}$$

Substituting Eqs. (A.9) and (A.8), and using the definitions of $\hat{C}_1$, C_1, and $\tilde{C}_1$, it is found that the firm faces the problem

$$\max_{\hat{C}, \eta_1, \eta_2} E[U(\hat{C}, \tilde{C}_1, \tilde{C}_2)] = \max_{\hat{C}, \eta_1, \eta_2} E[S(\hat{C})$$

$$+ [2(\hat{C} - C) + \varepsilon_1 / \lambda + \lambda(\eta_1 - \eta_1^*(\hat{C}))](\psi'(\eta_1^*(\hat{C}))$$

$$- 2\psi'(\eta_2^*(\hat{C}))) + [2(\hat{C} - C) + \varepsilon_2 / (1 - \lambda) + (1 - \lambda)(\eta_2 - \eta_2^*(\hat{C}))$$

$$+ 2\lambda(\eta_1 - \eta_1^*(\hat{C}))]\psi'(\eta_2^*(\hat{C})) - \lambda \psi(\eta_1) - (1 - \lambda)\psi(\eta_2)]. \tag{A.10}$$

The first-order conditions for η_1 and η_2 establish that

---

[26]Briefly, Baron and Besanko show that $U''(C) < 0$ violates local incentive compatibility.

$$\eta_i = \overset{*}{\eta_i}(\hat{C}), \quad i = 1, 2. \tag{A.11}$$

Use the firm's first-order condition, from Eq. (A.10), with respect to $\hat{C}$ and substitute the definition of $S'(\hat{C})$ to obtain

$$\hat{C} = C$$

for $C \leq C_z$. For $C > C_z$, use $S'(\hat{C}) = 0$, $\overset{*}{m_1}(\hat{C}) = 1/\lambda$, $\overset{*}{m_2}(\hat{C}) = -1/(1-\lambda)$, and $[\partial \overset{*}{\eta_i}(\hat{C})/\partial \hat{C}] = 0$ to obtain $\hat{C} = C$. The second-order conditions require that the Hessian (in this case, a 3-by-3 matrix) of second derivatives of $U(\hat{C}, \bar{C}_1, \bar{C}_2)$ (taken with respect to $\hat{C}$, η_1, and η_2) is negative semidefinite at the optimum. The Hessian for $C \leq C_z$ is

$$J = \begin{bmatrix} 2\psi_1'' \eta_1' \left(1 - \dfrac{\lambda}{2} \eta_1'\right) - 2\psi_2'' \eta_2' \left(1 - \dfrac{1-\lambda}{2} \eta_2'\right) & \lambda \psi_1'' \eta_1' & (1-\lambda)\psi_2'' \eta_2' \\ \lambda \psi_1'' \eta_1' & -\lambda \psi_1'' & 0 \\ (1-\lambda)\psi_2'' \eta_2' & 0 & -(1-\lambda)\psi_2'' \end{bmatrix}.$$

It can be verified that J is negative semidefinite at $\hat{C} = C$. For $C \geq C_z$, the Hessian, denoted J', becomes

$$J' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\lambda \psi_1'' & 0 \\ 0 & 0 & -(1-\lambda)\psi_2'' \end{bmatrix},$$

which is negative semidefinite.

Proof of Proposition 6:

(i) Forming the Hamiltonian of P2 and taking the derivatives with respect to η_i yields

$$\overset{*}{\eta_1}(C) = \frac{2-\lambda}{2K} - \frac{\alpha}{1+\alpha} \frac{2}{\lambda} H(C) + \frac{1}{1+\alpha} \frac{2}{\lambda} \frac{1}{f(C)} \int_{C_{min}}^{C} \gamma(\upsilon) \, d\upsilon \tag{A.12}$$

and

$$\overset{*}{\eta_2}(C) = \frac{1-\lambda}{2K}.$$

For $U(C) > 0$, then $\gamma(C) = 0$ and the last term in Eq. (A.12) is zero, which gives $\overset{*}{\eta_1}(C)$. Using (ICL), $U'(C) = -2K\eta_1(C) + 1 - \lambda = -1 + [\alpha/(1+\alpha)](4K/\lambda)H(C)$. Differentiating $U'(\hat{C} \mid C)$ at $\hat{C} = C$ yields $U''(C) = [\alpha 4K/(1+\alpha)\lambda]H'(C) > 0$.

(ii) Setting $U'(C_y) = 0$ gives $\eta_1^*(C_y) = [(1-\lambda)/2K]$. The proof of part (ii) is similar to the proof of part (ii) of Proposition 4 and is omitted.

(iii) Denote profit under full commitment as $U_f(C)$ and profit under renegotiation as $U_r(C)$. Then

$$U_r(C) = -\int_C^{C_y}\left(-1 + \frac{\alpha}{1+\alpha}\frac{4K}{\lambda}H(C)\right)\,dC$$

$$U_f(C) = -\int_C^{C_z}\left(-1 + \frac{\alpha}{1+\alpha}\frac{4K}{1-\lambda}H(C)\right)\,dC.$$

Also note that $C_z < C_y$ for all $\lambda > 0$. Subtracting gives

$$U_r(C) - U_f(C) = \begin{cases} \int_C^{C_z}\frac{\alpha}{1+\alpha}\frac{4K}{\lambda}H(C)\left(\frac{1}{1-\lambda}-1\right)\,dC + U_r(C_z) & \text{if} & C < C_z \\ U_r(C) & \text{if} & C_z \le C < C_y \\ 0 & \text{if} & C \ge C_y. \end{cases}$$

This is nonnegative.

INCENTIVES, THE BUDGETARY PROCESS, AND INEFFICIENTLY LOW PRODUCTION RATES IN DEFENSE PROCUREMENT
by William P. Rogerson[1]

1. INTRODUCTION

A widely held view in the defense community is that production of weapons systematically occurs in plants that are too large relative to the outputs actually produced.[2] An important consequence of this phenomenon is that existing outputs could therefore be produced more cheaply in smaller plants.[3] The traditional explanation for this phenomenon seems to be that Congress or the military continually and persistently overestimate the rate at which procurement will occur and thus always make the mistake of building production lines that are too big. This theory presumes that planners are too incompetent to ever learn from their past mistakes. Such irrationality seems implausible. This paper suggests an alternative theory—one that does not rely on irrational actors. It will be argued that the organization of the decisionmaking process itself creates incentives for rational actors to strategically choose too high a scale for plants.

The theory of this paper is based on the fact that Congress does not itself make technical design choices over production facilities; lacking sufficient time and expertise, it delegates these decisions to the military. Congress, of course, decides how many units of each weapon system to buy. Therefore, on a formal level, the decisionmaking process leading up to the production of a new weapon system is modelled as a two-person game where the military first chooses scale and then Congress decides what quantity to purchase (if any). The military's goal is to maximize the number of units produced. The key idea is that by increasing scale, the military lowers marginal

[2]See Air Force Systems Command (1983), Congressional Budget Office (1987), Gansler (1989), and Rogerson (1989).

[3]Because of the need for surge capacity (i.e., extra capacity to be used in the event of war), it may be that the optimal scale is somewhat larger than that which would minimize production cost of the planned peacetime rate. The widely held view is that capacity is too large even considering this factor.

cost and thus increases the amount that Congress will buy (so long as it buys any). Thus, the military will expand scale until production is so inefficient that Congress is indifferent between buying and not buying the system. This maximizes the number of units purchased.

In the resulting equilibrium, more output is produced than Congress would ideally want. However, there is another problem as well. The military cannot simply order Congress to increase output until all social surplus vanishes. Rather, it must manipulate the production technology so that Congress will rationally want to order more units. In the resulting equilibrium, the scale of production is too large given output. That is, the same output could be produced more cheaply using a smaller plant. Thus, inefficient production is essentially an unintended by-product of the military's attempts to expropriate the social surplus arising from weapon programs by inducing Congress to increase quantities purchased.

The model of this paper also suggests that the military may have the incentive to purposefully avoid using technologies that can flexibly produce at a variety of output levels with little change in average cost. The reason this may be so is very simple. The idea underlying the model is that the military can force Congress to procure a larger number of units by precommitting to a technology that penalizes Congress for procuring low quantities and rewards it for procuring high quantities. Very flexible technologies would frustrate this ability. This theory may help explain the slow rate of adoption of computerized flexible manufacturing technology in defense procurement.

Perhaps the most important policy prescription to flow from this analysis is simply to emphasize the fact that Congress cannot rely on the military to choose a cost-minimizing production technology even if the military's goal is to maximize military preparedness and Congress and the military agree on what constitutes military preparedness. Therefore, Congress needs to critically evaluate whether correct capacity choices are being made. Of course, direct monitoring can never provide a perfect solution. As stated above, decisions regarding the configuration of production facilities are delegated to the military to some extent because Congress has neither the time nor expertise to make these decisions. Nonetheless, Congress does perform some oversight and, naturally enough, concentrates its efforts on areas where it suspects the military will have incentives to make choices other than those Congress would make itself. The point of this paper is that scale choice is such an area.

This paper's model is most closely related to another (Rogerson, 1990) in which it is shown that a similar type of model can explain the apparent bias of procurement policy toward quality and away from quantity.[4] In that paper, quality plays the same type of role as scale does in this paper, i.e., the military chooses quality first and then Congress chooses quantity second. The chief technical difference between the models is that scale is not an attribute of military end-products whereas quality is. Therefore, in the current paper, only quantity directly affects military preparedness;

---

[4]Also see Lewis (1986), which considers a model of congressional-military interactions. Lewis (1986) is focused on different issues from the model of this paper.

in the previous paper, both quantity and quality do so. This simpler environment allows the derivation of two stronger conclusions than was possible before. First, in equilibrium, scale choice is always too high. (In the previous paper, one could only characterize the cases where quality choice was always too high or too low.) Second, in equilibrium, Congress earns zero surplus, i.e., scale is expanded until Congress is indifferent between purchasing and not purchasing the weapon. (In the previous paper, Congress generally strictly prefers to purchase the weapon in equilibrium.) On an economic level, the marginal contribution of this paper over Rogerson (1990) is that it identifies a different type of decision variable, which might be distorted because of the military's incentive to take advantage of first-mover opportunities. Scale choice is an important and much-debated aspect of military procurement and, thus, theories explaining distortions in this decision are potentially important. It may also be that predictions regarding scale choice may be more empirically testable than predictions regarding quality choice.

The paper is organized as follows. Section 2 presents the model and then Section 3 derives the major analytic results. Section 4 discusses the role of the key economic assumption of this paper that increasing scale lowers marginal cost. Section 5 analyzes the effect of fixed budget levels on the incentives of the military. Sections 6 and 7 apply the model to consider the issue of flexible production technologies.

2. THE MODEL

Assumptions and Notation

A weapon program will be completely described by two nonnegative numbers (x,s), where x denotes the number of units procured and s denotes the scale of production. The social value (in dollars) of x units of the weapon is given by $V(x)$. Both Congress and the military agree on $V(x)$. The cost of producing x units given the scale s is denoted by $C(x,s)$. For any fixed scale, C will be assumed to be a regular well-behaved cost function. Thus, each s represents a choice of a particular production technology and it is assumed that a continuum of technologies is available indexed by $s \in (0,\infty)$. As will be explained below, an assumption on $C(x,s)$ will be made that allows one to naturally interpret s as a measure of scale of production.

The above formulation of the problem abstracts away from two features of the real world that would complicate the analysis without altering the basic conclusions of the paper. First, in reality, production of a major weapon system occurs over a number of years and Congress must therefore choose both an annual rate of production and the total number of years that production will occur. In the model of this paper, all production is formally assumed to occur in a single period. However, it is straightforward to see that the model can also be interpreted as one where Congress chooses an annual rate of production, x, but the total number of years of production remains at some fixed level, n. In an earlier version of this paper (Rogerson, 1989), it is shown that all of the conclusions of this paper generalize in a straightforward fashion to the case where Congress chooses both x and n. Second, in reality both V and C are likely to be known only probabilistically at the time of the scale decision. In the model of this paper, both functions are assumed to be known with certainty.

Once again, the results of this paper can also be derived in more complicated models that allow for uncertainty.[5]

The various regularity assumptions that will be made about V and C will now be presented and discussed where necessary. For the most part, they simply guarantee that various maximization problems have unique well-behaved solutions. Assumptions 1–3 state the smoothness and concavity properties of V and C.

Assumption 1:

   (i) $V(x)$ is twice continuously differentiable over $(0,\infty)$

   (ii) $V(x)$ is strictly increasing over $(0,\infty)$

   (iii) $V(0) = 0$.

Assumption 2:

   (i) $C(x,s)$ is twice continuously differentiable over $(0,\infty)^2$

   (ii) C is strictly increasing in x over $(0,\infty)$ for every s

   (iii) $C(0,0) = 0$

   (iv) $C(x,s) > 0$ for $(x,s) \neq (0,0)$.

Assumption 3: $V(x) - C(s,x)$ is globally strictly concave over $(0,\infty)^2$.

A first-best program maximizes social surplus. It is formally defined as follows.

Definition: A weapon program is first best if it solves the following program.

$$\text{Maximize}\quad V(x) - C(x, s)$$
$$x, s \geq 0. \tag{1}$$

It will be assumed that a unique first-best program exists and that it is strictly preferred to producing nothing.

Assumption 4: A unique first-best weapons program denoted by (x^*,s^*) exists. Furthermore,

$$V(x^*) - C(x^*,s^*) > 0. \tag{2}$$

A scale is second best given an output if it minimizes the costs of producing that output.

---

[5]Note that one would formally model the need for surge capacity by allowing V to be random. That is, the function V would depend on what type of war (if any) is occurring and this would be known only probabilistically at the time of scale choice.

Definition: A scale is second best given x if it solves the following problem:

$$\text{Minimize} \quad C(x, s)$$
$$s \geq 0. \tag{3}$$

It will be assumed that a unique second-best scale exists for every $x > 0$.

Assumption 5: A unique second-best scale given x exists for every $x > 0$. Let $\phi(x)$ denote this value. Furthermore $\phi(x) > 0$ for all $x > 0$. Note that by Assumption 2, parts (iii) and (iv), $s = 0$ is the second-best scale when $x = 0$.

The short-run cost curve for a given scale $\hat{s}$ is simply $C(x, \hat{s})$. By Assumptions 2, part (iv), and 5, the long-run cost curve is also well defined. Let $L(x)$ denote the long-run cost curve. It is given by

$$L(x) = \begin{cases} 0 & , & x = 0 \\ C(x, \phi(x)), & x > 0. \end{cases} \tag{4}$$

The above assumptions guarantee that $L(x)$ is smooth over $(0, \infty)$. However, it may jump at zero.

It will be useful to define the second-best output over the nonnegative and positive orthants. These will be called, respectively, the second-best and interior second-best outputs.

Definition: An output is second best (interior second best) given s if it maximizes the following function over $x \geq 0$ $(x > 0)$.

$$V(x) - C(x, s). \tag{5}$$

Assumption 6: A unique interior second-best output given s exists for every $s > 0$. Let $\psi(s)$ denote this value.

It will also be assumed that there exists a capacity level $\bar{s}$ such that the optimal interior second-best output yields negative social welfare for values of $s > \bar{s}$. This simply means that large enough capacity levels would result in it being *ex ante* preferable to produce nothing. (If one had to build a plant able to produce one billion F-16s then it might be better not to build the plant and have none.) For expositional convenience, it will also be assumed that the interior maximum produces strictly positive social welfare for values of s less than $\bar{s}$. However, this is not actually necessary for the results.

Assumption 7: There exists an $\bar{s} > 0$ with the following property:

$$V(\psi(s)) - C(\psi(s), s) \gtreqless 0 \iff s \lesseqgtr \bar{s}. \tag{6}$$

Let $\bar{x}$ denote the interior second-best output at $\bar{s}$, i.e.,

$$\bar{x} = \psi(\bar{s}).\tag{7}$$

This completes the technical regularity assumptions. Now the assumption that allows one to interpret s as a "scale" parameter will be presented and explained.

Assumption 8:

$$C_{xs}(x,s) < 0.\tag{8}$$

According to Assumption 8, increasing the scale of production lowers the marginal cost of production.[6] This is intuitively reasonable. Intuitively, a large-scale plant is one where greater fixed costs are incurred to reduce marginal costs. Support for the contention that this is a reasonable definition of scale can be found by examining how the second-best scale choice varies with x. Total differentiation of the first-order condition for cost-minimization yields

$$\phi'(x) = \frac{C_{sx}}{-C_{ss}}.\tag{9}$$

Since the denominator must be positive by the concavity assumption (Assumption 3), it can be seen that a larger scale is optimally chosen for a larger output if and only if C_{sx} is positive.

Equilibrium

The decisionmaking process leading up to the adoption of a new weapon program will now be described. Three fundamental assumptions about the nature of this process will be made.

First, it will be assumed that the military considers military value only when deciding between two programs and ignores cost. That is, the military prefers one weapons program (x,s) to another $(\hat{x}, \hat{s})$ if and only if x is greater than $\hat{x}$.

Second, it will be assumed that the decisionmaking process is sequenced as follows. First, the military chooses a scale of production. Then Congress chooses a level of production taking the scale decision as given. This assumption reflects the fact that the military has greater technical expertise than Congress. Thus, Congress must delegate the determination of a production technology to the military. Given the technology selected by the military, Congress can calculate the costs of producing various levels of output. By choosing a funding level for the program, it then determines how many units will be produced. However, it is incapable of determining whether other production technologies might have resulted in lower costs.

---

[6]For expositional simplicity, it is assumed that this holds globally. It will be clear in Section 3 that all that is required is that this hold over the one-dimensional manifold defined by $\{(x, s): x = \psi(s)$ and $s^* \geq s \geq \bar{s}\}$.

Obviously, both of these assumptions are somewhat extreme. In reality, military planners may care about more than the success of their program and Congress may be able to evaluate some aspects of the choice of production technology. Nonetheless, it certainly seems plausible that planners care more about the success of their programs than whether social value is maximized and that Congress has less expertise than the military regarding the effects of plant scale on production cost. This paper makes extreme versions of both of these assumptions to clearly illustrate their effects in the simplest possible model.

Third, it will be assumed that the defense contractor (who is perfectly controlled by the military) initially pays for the production facility when it is built. Then Congress pays for the facility as part of the production cost if and only if it purchases any units of the system. That is, Congress has the option of purchasing zero units and paying zero dollars. However, it pays for the entire production facility if it purchases any units.

There are two reasons for making this assumption. First, this is to a large extent the way that the procurement process is organized. Second, if it was assumed that Congress always paid for facilities capital investments authorized by the military regardless of eventual production decisions, there would be another more obvious incentive for the military to increase capacity above the first best. The basic idea is that Congress will ignore facilities capital costs when making its adoption decision if it has already paid these costs. Thus, the military can increase the probability that Congress will adopt a program by decreasing variable costs through employing greater amounts of sunk facilities capital expense. The incentive for the military to choose excess capacity identified by this paper is totally separate and unrelated to this. The clearest method of demonstrating this is to simply assume that Congress follows the policy of paying zero if it does not purchase any units.

The above three assumptions result in the following structure to the game between Congress and the military. As usual, it is most convenient to work backward. Given the military's choice of s, the Congress will then choose a value of x to maximize

$$V(x) - C(x,s) \tag{10}$$

so long as it can achieve greater than zero surplus. Otherwise it will choose $x = 0$. Let $\xi(s)$ be the correspondence denoting Congress's choice given s. By the assumptions in Part A above, ξ is single valued for all values of s except $\bar{s}$. For smaller values, it is the interior second-best output and for larger values it is zero. For $\bar{s}$, both $\bar{x}$ and 0 are optimal. Formally

$$\xi(s) = \begin{cases} \psi(s) & , \quad s < \bar{s} \\ \{0, \ \bar{x}\} & , \quad s = \bar{s} \\ 0 & , \quad s > \bar{s}. \end{cases} \tag{11}$$

At the beginning of the game, the military therefore chooses s to maximize $V(x)$ realizing that s affects x as described above.

Formally, then, an equilibrium weapon program is described as follows.

Definition: An equilibrium weapon program solves the following problem.

$$\text{Maximize}_{x, s} \quad V(x) \tag{12}$$

$$\text{subject to } x \in \xi(s) \tag{13}$$

3. FORMAL ANALYSIS

Since V is strictly increasing in x, the military's problem boils down to choosing s to maximize $\xi(s)$. Thus, to describe capacity choice, one needs to describe the behavior of $\xi(s)$. Recall that $\xi(s)$ equals the interior solution, $\psi(s)$, for $s \leq \bar{s}$ and equals zero for $s \geq \bar{s}$. Proposition 1 describes the critical feature of $\psi(s)$ for the purposes of this paper. Namely, it is strictly increasing. The reason for this is very simple and intuitive. Increases in scale result in lower marginal costs. This results in a larger interior maximum.

Proposition 1:

$$\psi'(s) > 0. \tag{14}$$

Proof: The interior maximum for any s is determined by the first-order condition

$$V'(x) = C_x(x, s). \tag{15}$$

Total differentiation yields

$$\frac{dx}{ds} = \frac{C_{xs}(x, s)}{V''(x) - C_{xx}(x, s)}. \tag{16}$$

This is positive by Assumptions 3 and 8.

$$\text{QED.}$$

The nature of the equilibrium program is now clear. The military can induce Congress to buy more units by choosing a larger scale of production and thereby lowering marginal cost. However, production at scales larger than $\bar{s}$ is so inefficient that Congress would rather cancel the program. Thus, the optimal course of action for the military is to increase scale to this point of indifference. This is stated as Proposition 2.

Proposition 2: The unique equilibrium weapon program is $(\bar{x}, \bar{s})$.

Proof: As above.

<div align="center">QED.</div>

The welfare properties of the equilibrium program can now be analyzed. This is done in Proposition 3.

Proposition 3:

(i) The equilibrium capacity and output are both strictly greater than the first best, i.e.,

$$\bar{x} > x^* \tag{17}$$

and

$$\bar{s} > s^*. \tag{18}$$

(ii) The equilibrium output is second best given the equilibrium capacity, i.e.,

$$\bar{x} = \psi(\bar{s}) \tag{19}$$

and

$$V(\bar{x}) - C(\bar{x}, \bar{s}) \geq V(0) - C(0, \bar{s}). \tag{20}$$

(iii) The equilibrium capacity is strictly greater than the second-best capacity given output, i.e.,

$$\bar{s} > \phi(\bar{x}) \tag{21}$$

Proof: Parts (i) and (ii) follow immediately. Part (iii) is an immediate consequence of the concavity of the problem. The general result on which this depends is stated and proven in the appendix.

<div align="center">QED.</div>

These results can be very clearly illustrated on a graph of average costs. Let $AC(x,s)$ denote the average cost of producing x units given scale s. Let $AL(x)$ denote the average long-run cost curve. Now refer to Figure 9.1. The long-run average cost curve is shown to be declining, since this is probably true over the range in which most weapons would be purchased. The quantity x^* is the optimal quantity. The capacity s^* is the optimal capacity at which to produce x^*. This means that $AC(x,s^*)$ is just tangent to $AL(x)$ at x^* as drawn. The quantity $\bar{x}$ is actually purchased and this is greater
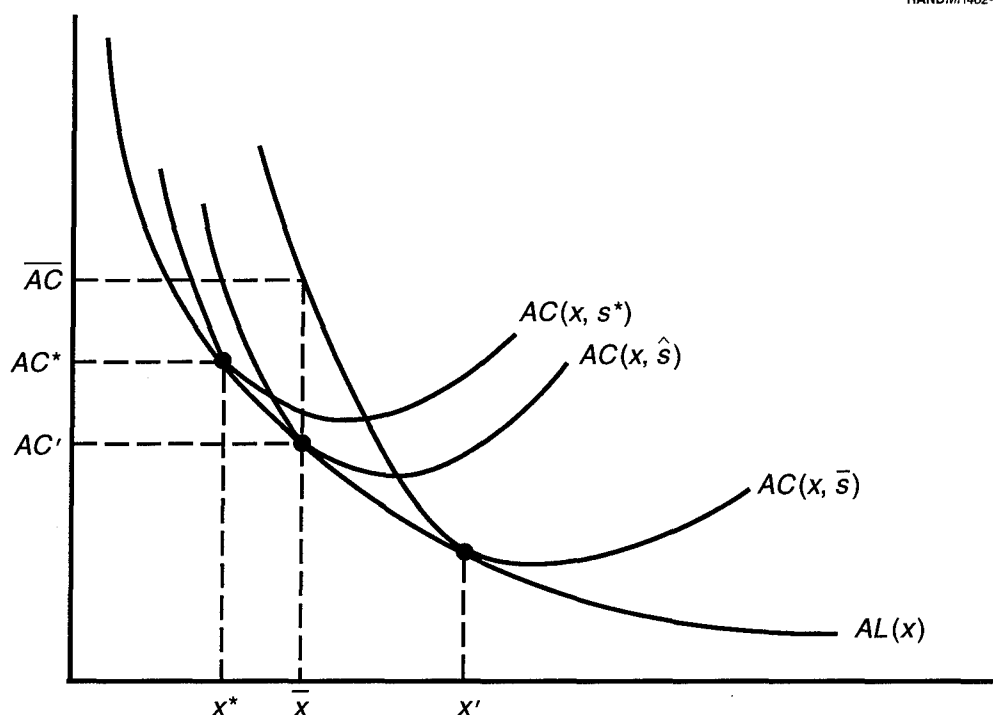
RANDMR462-9.1

Figure 9.1—The Equilibrium Weapons Program

than x*. Let s' denote the optimal scale technology at which to produce $\bar{x}$. Therefore, $AC(x, s')$ is drawn to be tangent to $AL(x)$ at $\bar{x}$. However, this is *not* the technology that is used in equilibrium. Rather, a larger-scale technology $\bar{s}$ is used. Therefore, $AC(x, \bar{s})$ is drawn to be tangent to $AL(x)$ at a point to the right of $\bar{x}$, which is labelled x'.

The first-best program would have been produced at an average cost of AC^*. If the actual quantity chosen was produced efficiently, it would be produced at a lower average cost, AC'. Even this would have been a socially inferior choice, because the marginal costs would exceed the marginal benefits. However, the actual outcome is even worse, because $\bar{x}$ is produced using an inefficiently large technology. This results in average costs of $\overline{A}\ \overline{C}$. ($\overline{A}\ \overline{C}$ may or may not be larger than AC^*.)

Therefore, there are two inefficiencies in equilibrium. First, the quantity produced is too large even if it were produced efficiently. Second, an efficiently large-scale production technology is used.

To conclude this section, this theory's explanation for why military planners seem to persistently overestimate future output levels for new weapon systems will be explicitly drawn out. Suppose that a military planner wants to induce Congress to increase the output of his program, just as modelled in this paper. Suppose that the socially optimal quantity is 50. However, the planner knows that if he can build a plant designed to produce 100, he can induce Congress to purchase 75. How does the planner accomplish his goal? The solution is to announce a projected procure-

ment of 100 and to instruct all technical personnel involved to design a plant that will efficiently produce the projected rate. Once the plant is built, Congress is induced to purchase 75 just as the military planner intended. That is, if projected output rates are instructions for scale decisions, then the theory of this paper predicts that projected rates will always exceed actual rates. This is just another way of saying that scale is too large in equilibrium.

4. THE EFFECT OF SCALE ON MARGINAL COST

This section will discuss the role played by Assumption 8, that increasing scale lowers marginal cost, in generating the results of this paper. As explained in Section 2, the assumption that C_{sx} is positive is essentially definitional and thus relatively noncontroversial. However, this is not the end of the story. It will be shown below that the existence of significant production inefficiencies requires that scale choice have a significant effect on marginal cost. In particular, as C_{sx} goes to zero, the theorems of this paper remain true, but the size of the production inefficiency goes to zero.

This point can be illustrated most clearly by considering the case where C_{sx} equals zero. Assume that there are only two inputs, labor and capital, and interpret capital as the scale choice. Assume that capital and labor must be used in fixed proportions, i.e., there is no substitutability between capital and labor. Without loss of generality, assume that one unit of output requires one dollar of labor expenditure and one dollar of capital expenditure. Therefore, the cost function is given by

$$C(s, x) = \begin{cases} x, & x \le s \\ \infty, & x > s. \end{cases} \tag{22}$$

The nature of equilibrium for this case is illustrated in Figure 9.2. The line with a slope of 1, labelled ℓ_1, represents the short-run marginal cost of production assuming that sufficient capital is available. The line with a slope of 2, labelled ℓ_2, represents the long-run marginal cost, i.e., the cost of the labor and capital. The downward sloping line is the marginal benefit of output. The socially optimal output is, of course, x^*, where long-run marginal cost equals marginal benefit. Let x^{**} denote the output where short-run marginal cost equals marginal benefit. The equilibrium output in this model is determined as follows. If the total benefit is greater than the total cost at x^{**}, then the military will select x^{**} units of capital and Congress will choose to produce x^{**} units of output. If this condition is not satisfied, then define $\hat{x}$ to be the largest output such that total benefit equals total cost in the interval (x^*, x^{**}). The military will select $\hat{x}$ units of capital and Congress will choose to produce $\hat{x}$ units of output.

The important point to note is that in either case, there is no excess scale in equilibrium. Thus, the military induces Congress to purchase more output than the first-best level, but there is no production inefficiency. The explanation for this is very intuitive. As stated in the Introduction, production inefficiency is an unintended by-product of the military's efforts to induce greater production. Congress can usually respond to excessive levels of scale by producing less and using some of the excess
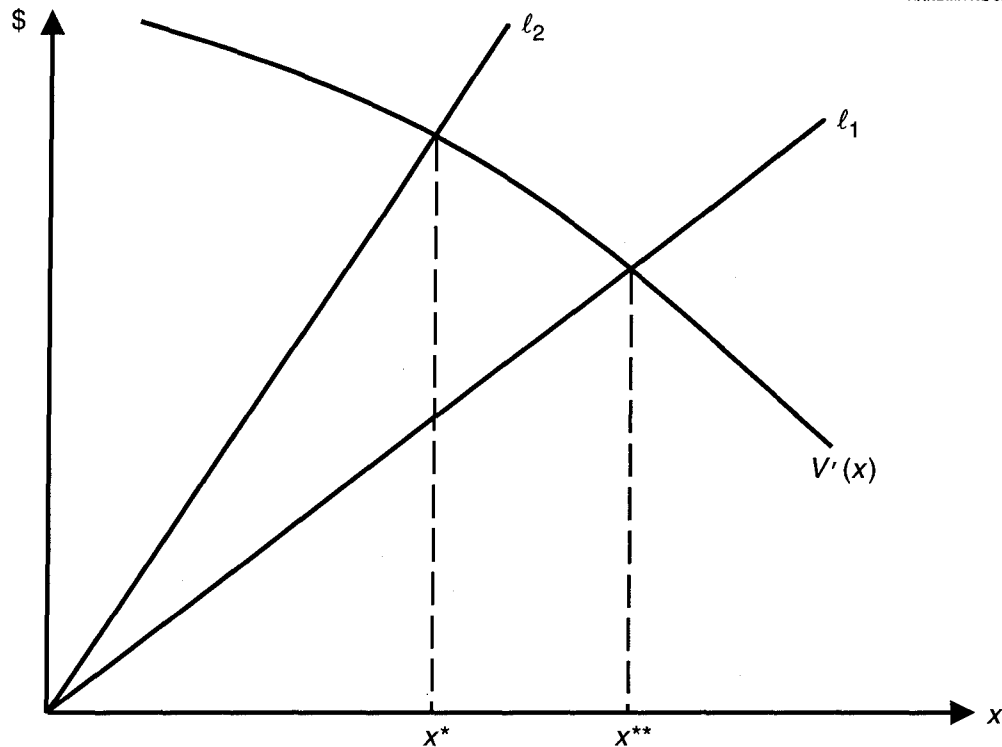
Figure 9.2—Equilibrium with Fixed Input Proportions

fixed inputs as substitutes for the variable inputs. However, in the example it was assumed that absolutely no substitution was possible. This is why there is no productive inefficiency.

Whether scale choice has a significant effect on the marginal cost of defense production is an empirical question beyond the scope of this paper. There is no published work that sheds much light on this issue. A number of studies, such as Hildebrandt and Sze (1986), estimate cost functions for products but these analyses do not separately identify inputs and instead implicitly assume that production occurs on the long-run cost curve.

If one could observe the nature of production in defense plants, the following three qualitative factors would be relevant to determining if increased scale lowers marginal cost. The first factor is whether production in defense plants designed to produce high rates of output is organized in a significantly different fashion from production in defense plants designed to produce low rates of output. For example, it may be that plants designed for high-rate production are more automated and involve a larger number of more specialized job tasks. If this is true, it seems likely that larger plants would exhibit lower marginal costs when production rates were significantly lower than those planned for. This is because increased automation and specialization of job tasks raises fixed costs and lowers marginal costs. One possible approach to gathering evidence on this factor would be to compare the production of

missiles and aircraft, since missiles generally are produced at much higher rates. Knowledgeable industry participants state that missile production is significantly more automated than aircraft production. However, I am not aware of any empirical analysis on this question.

The second qualitative factor is whether plants with excess capacity appear to be using all of their machinery and capital or whether a portion of it is sitting perfectly idle. If the capital in plants with excess capacity is being used as part of the production process, this suggests that it is substituting for variable inputs and thus lowering the marginal cost of production. However, capital that is perfectly idle may not be lowering the marginal cost of production.

Third, even capital investments that appear to be used in fixed proportions with labor may lower marginal cost because of reducing the need for multiple shifts. Defense plants can operate, and sometimes are operated, using two or even three shifts of labor. However, the marginal cost of adding a second and particularly a third shift can be much higher than the marginal cost of the first shift. This is because of the need for nonstandard work hours and because maintenance and set-ups become more difficult to schedule. Thus, even if labor and capital appear to be used in fixed proportions in a given shift, it may be that expanding capacity so that all production can occur in one shift may lower marginal cost.

In summary, it is not clear *a priori* whether scale expansion of defense plants lowers marginal costs significantly. The incentive effects of this paper will be significant only to the extent that this is true. Although it seems plausible that this may be true, a definitive answer awaits empirical analysis.

5. FIXED BUDGETS

Suppose that Congress was able to precommit to a fixed budget level for a program regardless of the scale chosen by the military. In this case, the military would choose the second-best program given the budget level.[7] Therefore, if Congress were to precommit to a budget exactly sufficient to fund the first-best weapons program, the military would choose the first-best program.

The above argument makes it tempting to view precommitment to fixed program budgets as offering a complete solution to the problem of provision of excess capacity. There are a number of problems with this solution, however. First, to calculate the first-best budget, Congress must in general know the entire function $C(x,s)$. In this case, Congress would not need to delegate decisionmaking authority to the military. It could simply instruct the military to choose s^*. Second, for a major weapon system, procurement will occur over a ten- or even twenty-year period. It is hard to

---

[7]Formally, a second-best program given the budget level B is a solution to the following problem

$$\text{Maximize}_{x,\, s} \quad V(x, s)$$

$$\text{Subject to} \quad C(x, s) \le B.$$

believe that Congress could precommit ten years in advance to anything. Furthermore, many factors in the environment will change between the time capacity is chosen and the quantity decisions are made. This means that Congress would in reality have to precommit to a budget *rule*, i.e., a rule describing what the budget will be each year as a function of the environmental factors. However, describing the set of all contingencies in an objectively verifiable fashion and the budget level for each one would probably be an impossible task.[8]

Thus, precommitment to fixed budget levels clearly does not provide a complete solution. Nonetheless, this point is still interesting for a number of reasons. First, it highlights the two key problems that any proposed solution must face—Congress's inability to precommit and Congress's lack of information and expertise, especially at the planning stage. Second, the fact that choosing a fixed budget does not generate the first-best outcome when Congress is not perfectly informed does not mean it should not be used if precommitment was possible. After all, the previous sections show that the alternative of waiting for the military to precommit to a capacity choice does not yield a first-best outcome either. This is an interesting question for future research. Third, it suggests that the incentive problems identified in this paper will be most severe in situations where exogenous factors have not fixed the available budget. For example, it may be that central planners within the DoD may view the total defense budget as relatively fixed. In this case, expansion of the scale on all programs may not be an attractive strategy to them. However, individual services or advocates of particular programs may view their potential share of the budget as highly variable and respond by purposefully increasing scale. Thus, it may be that excessive scale is the result of a competition between services and programs for greater budget shares, rather than the result of the military as a whole attempting to increase defense spending.

6. INCENTIVES TO CHOOSE INFLEXIBLE TECHNOLOGIES

On an intuitive level, one production technology can be thought of as exhibiting more rate flexibility than another one if production is relatively efficient over a greater range of output rates. The purpose of this section is to argue that the same type of model as in the previous sections can be used to show that the military may well have an incentive to purposely avoid flexible technologies. The reason is very simple. The idea underlying the model of the previous sections is that the military can force Congress to procure a larger number of units by precommitting to a technology that penalizes Congress for procuring low quantities and rewards it for procuring high quantities. Very flexible technologies would frustrate this ability.

This idea can be very clearly illustrated as follows. Assume that the problem is well behaved as modelled in Section 2. This is illustrated in Figure 9.3. The curve $V(x)$ denotes the value of x units and $L(x)$ is the long-run cost curve. The curve $B(x)$ is the cost that would be incurred to produce x units if the military chose a capacity that induced Congress to select x. Formally $B(x)$ is defined by
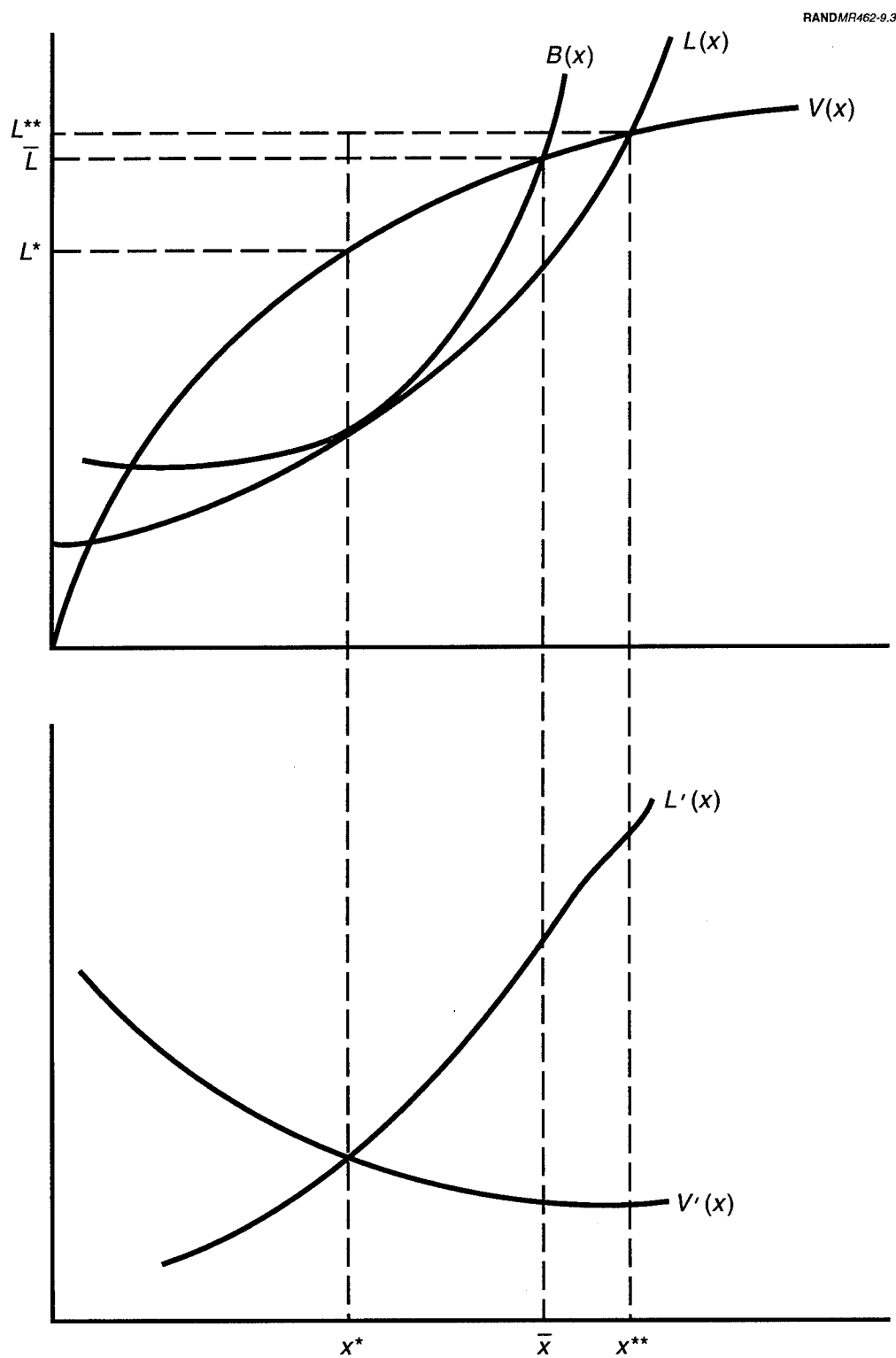
---

[8]See Rogerson (1990) for a formal model of this idea.

Figure 9.3—Incentives to Adopt Tailor-Made Inflexibilities

$$B(x) = C(x, \psi^{-1}(x)). \tag{23}$$

The first-best output is x^*, where marginal cost equals marginal benefit. If the military chooses the first-best capacity, s^*, then Congress will choose the first-best output, x^*. Therefore

$$B(x^*) = L(x^*). \tag{24}$$

However, to induce Congress to choose higher outputs, the military must choose higher capacities. Furthermore, as was shown in Proposition 3, the resulting weapon program will have a capacity greater than the second-best capacity given the output. Therefore, B is above L to the right of x^*. A similar argument shows that B is also above L to the left of x^*. Therefore, B is tangent to L at x^* as drawn.

The equilibrium output is $\bar{x}$. This is the highest output that Congress can be induced to select, because higher outputs would result in costs exceeding benefits and the entire program would be cancelled. Note in particular that it would be possible to produce greater outputs and have total costs be less than total benefits. In Figure 9.3 the point x^{**} is the largest such output. However, these points are not attainable because the equilibrium weapons program involves inefficient production, i.e., production off the long-run cost curve. Thus, given the inefficient production, $\bar{x}$ is the greatest attainable output.

Now suppose that the military had the option of making any production technology more inflexible. Formally assume that the military can choose any cost function $D(x)$ to present to Congress so long as there exists an s such that

$$D(x) \geq C(x,s) \tag{25}$$

for every x. That is, the military selects a scale just as before which determines the best attainable technology $C(x,s)$. However, it can now alter the technology to make costs of particular outputs rise if it wants to.

It is clear that the following choice is optimal for the military. Let s^{**} denote the scale of plant that is second-best given x^{**}. Then, the military will choose s^{**} and make the costs of selecting any $x < x^{**}$ prohibitively high. Formally, it will choose $D(x)$ as follows

$$D(x) = \begin{cases} \infty, & x < x^{**} \\ C(x, s^{**}), & x \geq x^{**}. \end{cases} \tag{26}$$

Faced with this cost function, Congress will choose x^{**} units of output. This must be optimal for the military because it can never induce an output higher than x^{**}.

This simple example illustrates the general point that certain types of rate inflexibilities can be of value to the military to the extent that they restrict Congress's choices in fashions desired by the military.

7. FLEXIBLE MANUFACTURING TECHNOLOGY

The automation revolution in manufacturing often referred to as flexible manufacturing technology (FMT) or computer-aided manufacturing has not yet arrived in defense production. Rather, what will be referred to as the standard manufacturing technology (SMT) is still used. In an extremely insightful article, two RAND researchers sum up the situation as follows:

> Today's defense manufacturing technology is still characterized by the kind of inflexible production line, pioneered by Henry Ford, that reached maturity in World War II. This production line is set up to produce a single design, in large quantities, over long periods of time. Although production lines have been progressively automated since 1960, the kind of automation adopted in the defense sector has done little to increase flexibility, and the procurement culture seems to have changed very little. Both the government buyer and the contractor seem to regard the specialized, optimized production line, designed for high rates of output, as the norm (Dews and Birkler, 1983, p. 1).

Dews and Birkler go on to describe two key features of the new technology in more detail. Both properties stem from the fact that the computerization of facilities allows unused capacity to be easily reprogrammed to produce some other product rather than to merely sit idle.

(i) It is much more efficient than the SMT for low production rates. In particular, they argue that production rates on most aerospace programs are low enough that using FMT would be more efficient.

(ii) Efficient production can occur over a relatively broad range of output rates.

In terms of the definitions introduced in this paper, Dews and Birkler claim that FMT is a lower-scale technology than SMT and that it is more flexible.

As the above quote indicates, they feel that no one, including the military, seems to be in much of a hurry to employ the new technology. They do not describe any reasons for this apparent reluctance other than perhaps institutional inertia. This paper, of course, supplies a theory that explains precisely why the military might be reluctant to adopt such a technology. According to this paper's theory, each of the characteristics (i) and (ii) supplies a possible reason for this reluctance. First, it may be the case that plants using FMT are of lower scale than those using SMT. That is, it may be efficient to use plants employing FMT for lower levels of production and plants using SMT for higher levels of production. Section 3 suggests that the military might well prefer to use SMT so long as Congress would still purchase the system, even though FMT would be more efficient. Second, it may be the case that FMT is capable of producing even large levels of output as efficiently as SMT. However, if FMT exhibits more rate flexibility in the sense that it would permit Congress to reduce the quantity purchased without significantly raising average cost, then the analysis of Section 6 suggests that the military may well prefer SMT.

The example of FMT also illustrates another important point. Namely, it is not the case that increases in scale are always synonymous with increases in capital inten-

sity. It could easily be the case that highly capitalized automated production facilities can operate efficiently at lower rates of output than can less highly capitalized facilities using older technology. Thus, the prediction of Section 3 that the military will choose inefficiently large scales of production is not necessarily a prediction that it will choose too high a capital intensity. In fact, if Dews and Birkler are correct, precisely the reverse may be true.

APPENDIX

Suppose that $F(x,y)$ is a twice continuously differentiable, strictly globally concave function. Assume that a unique value of y maximizes $F(x,y)$ for every x. Denote it by $\phi(y)$. Assume that a unique value of x maximizes $F(x,y)$ for every y. Call it $\psi(y)$. Finally assume that a unique maximum to $F(x,y)$ occurs at (x^*,y^*). Define $\gamma(y)$ by

$$\gamma(y) = \phi(\psi(y)). \tag{A.1}$$

The technical result required to prove Proposition 3.3(iii) is that

$$y \gtreqless y^* \iff y \lesseqgtr \gamma(y). \tag{A.2}$$

To prove this, note that y^* is the only value of y to satisfy

$$y = \gamma(y) \tag{A.3}$$

since there is a unique maximum. Therefore it is sufficient to prove that

$$\gamma'(y^*) < 1. \tag{A.4}$$

By total differentiation

$$\gamma'(y^*) = \frac{F_{xy}(x^*, y^*)^2}{F_{xx}(x^*, y^*)F_{yy}(x^*, y^*)}. \tag{A.5}$$

This is less than one by the strict concavity of F.

<div align="center">QED.</div>

Air Force Systems Command, *The Affordable Acquisition Approach,* Andrews Air Force Base, Maryland, 1983.

Alexander, A., *Weapons Acquisition in the Soviet Union, United States, and France,* RAND, P-4989, March 1973.

Antle, R., and A. Smith, "An Empirical Investigation into the Relative Performance Evaluation of Corporate Executives," *Journal of Accounting Research,* Vol. 24, Spring 1986, pp. 1–39.

Anton, J. J., and D. A. Yao, "Measuring the Effectiveness of Competition in Defense Procurement: A Survey of the Empirical Literature," *Journal of Policy Analysis and Management,* 1990.

Averch, H., and L. Johnson, "Behavior of the Firm Under Regulatory Constraint," *American Economic Review,* Vol. 52, 1962, pp.1052–1069.

Bailey, E., *Economic Theory of Regulatory Constraint,* Lexington Books, D. C. Heath and Co., Lexington, Massachusetts, 1973.

Baiman, S., and J. H. Evans, "Pre-Decision Information and Participative Management Control Systems," *Journal of Accounting Research,* Vol. 21, Autumn 1983, pp. 371–395.

Baron, D. P., "Regulatory Incentive Mechanisms and Political Action," Working Paper, Stanford Business School, Stanford University, 1988.

Baron, D. P., and D. Besanko, "Regulation, Asymmetric Information, and Auditing," *The RAND Journal of Economics,* Vol. 15, 1984, pp. 447–470.

Baron, D. P., and D. Besanko, "Commitment and Fairness in a Dynamic Regulatory Relationship," *Review of Economic Studies,* Vol. 54, 1987a, pp. 413–436.

Baron, D., and D. Besanko, "Monitoring Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting," *The RAND Journal of Economics,* Vol. 18, 1987b, pp. 509–532.

Baron, D., and D. Besanko, "Monitoring of Performance in Organizational Contracting: The Case of Defense Procurement," *Scandinavian Journal of Economics*, Vol. 90, 1988, pp. 329–356.

Baron, D., and R. Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica*, Vol. 50, 1982, pp. 911–930.

Bower, A. G., "Procurement Policy and Contracting Efficiency," *International Economic Review*, Vol. 34, 1993, pp. 873–901.

Bower, A. G., and K. Osband, "Defense Procurement Policy and Contracting Efficiency," mimeo, Stanford University, Stanford, California, 1990.

Brown, L., Palisades Generating Company/Consumers Power Company, prepared direct testimony before the Federal Energy Regulatory Commission, 1990.

Brown, L., M. Einhorn, and I. Vogelsang, *Incentive Regulation: A Research Report*, Office of Economic Policy, Federal Agency Regulatory Commission, Washington, D.C., 1989.

Bussing, I., *Public Utility Regulation and the So-Called Sliding Scale*, Columbia University Press, New York, 1936.

Caillaud, B., R. Guesnerie, P. Rey, and J. Tirole, "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions," *The RAND Journal of Economics*, Vol. 19, 1988, pp. 1–26.

Center for Research in Security Prices, *CRSP Stock Files Users Guide*, University of Chicago, Illinois, 1986.

Chamberlin, G., "Asset Pricing in Multiperiod Securities Markets," Harvard Institute of Economic Research, Discussion Paper No. 1401, 1988.

Che, Y.-K., "Design Competition, Cost-Based Regulation and Quality Bias: A Theory of Multi-Dimensional Auction and Its Applications," mimeo, Stanford University, Stanford, California, 1989.

Che, Y.-K., "Design Competition Through Multidimensional Auctions," *The RAND Journal of Economics*, Vol. 24, 1993, pp. 668–680.

Cheney, R., "Report to the President," *Defense Management Review*, Department of Defense, Washington, D.C., 1989.

Comptroller General of the United States, *Assessment of the Study of Defense Contractor Profitability*, GAD/NSIAD-87-50, 1986.

Congressional Budget Office, *Effects of Weapons Procurement Stretch-Outs on Costs and Schedules*, Washington, D.C., 1987.

Demski, J., D. Sappington, and P. Spiller, "Managing Supplier Switching," *The RAND Journal of Economics*, 1987, pp. 77–97.

Department of Defense, *Defense Financial and Investment Review*, 1985.

Dews, Edmund, and John Birkler, *Reform in Defense Acquisition Policies: A Different View,* RAND, P-6927, November 1983.

Dews, E., G. K. Smith, A. A. Barbour, E. D. Harris, and M. A. Hesse, *Acquisition Policy Effectiveness: Department of Defense Experience in the 1970s,* RAND, R-2516-DRE, October 1979.

Fama, E., "Risk-Adjusted Discount Rates and Capital Budgeting Under Uncertainty," *Journal of Financial Economics,* Vol. 5, 1977, pp. 3–24.

Fox, R. J., *Arming America: How the U.S. Buys Weapons,* Harvard University Press, Cambridge, Massachusetts, 1974.

Fox, R. J., *The Defense Management Challenge: Weapons Acquisition,* Harvard Business School Press, 1988.

Galbraith, J. K., "The Big Defense Firms Are Really Public Firms and Should Be Nationalized," *New York Times Magazine,* November 16, 1969.

Gansler, J. S., *The Defense Industry,* MIT Press, Cambridge, Massachusetts, 1980.

Gansler, J. S., *Affording Defense,* MIT Press, Cambridge, Massachusetts, 1989.

Gates, W., "Incentive Contracting and Principal-Agent Theory: Theory Is Nice, but Can It Be Applied?" mimeo, Naval Postgraduate School, Monterey, California, 1988.

General Accounting Office, "Incentive Contracts: Examination of Fixed-Price Incentive Contracts," briefing report to U.S. Senator Carl Levin, Washington, D.C., 1987.

Goldberg, Victor, "Regulation and Administered Contracts," *Bell Journal of Economics,* Vol. 7, Autumn 1976, pp. 426–448.

Gonik, J., "Tie Salesmen's Bonuses to Their Forecasts," *Harvard Business Review LVI,* May 1978, pp. 116–123.

Grossman, S., and O. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica,* Vol. 51, 1983, pp. 7–45.

Guesnerie, R., and J.-J. Laffont, "A Complete Solution to a Class of Principal Agent Problems," *Journal of Public Economics,* Vol. 25, 1984, pp. 329–369.

Hildebrandt, G., and M. Sze, *Accounting for the Cost of Tactical Aircraft,* RAND, N-2420-PAE, July 1986.

Holmström, B., "Moral Hazard and Observability," *Bell Journal of Economics,* Vol. 10, 1979, pp 74–91.

Huang, C., and R. H. Litzenberger, *Foundations for Financial Economics,* North-Holland, New York, 1988.

Johnson, T., and R. Kaplan, *Relevance Lost,* Harvard Business School Press, Cambridge, Massachusetts, 1987.

Kamien, M. I., and N. L. Schwartz, *Dynamic Optimization*, Elsevier Science Publishing Company, Amsterdam, 1981.

Kaplan, R. S., and A. Atkinson, *Advanced Management Accounting*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

Kirby, A., S. Reichelstein, P. K. Sen, and T. Paik, "Participation, Slack and Budget Based Performance Evaluation," *Journal of Accounting Research*, Vol. 29, Spring 1991, pp. 109–128.

Laffont, J.-J., and J. Tirole, "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, Vol. 94, 1986, pp. 614–641.

Laffont, J.-J., and J. Tirole, "Auctioning Incentive Contracts," *Journal of Political Economy*, Vol. 95, 1987, pp. 921–937.

Laffont, J.-J., and J. Tirole, "Repeated Auctions of Incentive Contracts, Investment, and Bidding Parity with an Application to Takeovers," *The RAND Journal of Economics*, Vol. 19, 1988a, pp. 516–537.

Laffont, J.-J., and J. Tirole, "Adverse Selection and Renegotiation in Procurement," Social Science Working Paper No. 665, California Institute of Technology, 1988b.

Laffont, J.-J., and J. Tirole, "The Dynamics of Incentive Contracts," *Econometrica*, Vol. 56, 1988c, pp. 1153–1175.

Laffont, J.-J., and J. Tirole, "A Theory of Incentives in Procurement and Regulation," mimeo, University of Toulouse, Toulouse, France, 1991.

Lambert, R., and D. Larcker, "An Analysis of Accounting and Market Measures in Executive Contracts," *Journal of Accounting Research*, Vol. 25 (supplement), 1987, pp. 85–125.

Lewis, Tracy, "Budget Competition in Project Funding," mimeo, University of California at Davis, 1986.

Lichtenberg, F. R., "The Private R&D Investment Response to Federal Design and Technical Competitions," *American Economic Review*, Vol. 78, 1988a, pp. 550–559.

Lichtenberg, F. R., "Government Subsidies to Private Military R&D Investment: DoD's IR&D Policy," mimeo, 1988b.

McAfee, R. P., and J. McMillan, "Bidding for Contracts: A Principal-Agent Analysis," *The RAND Journal of Economics*, Vol. 17, 1986, pp. 326–328.

McAfee, R. P., and J. McMillan, "Competition for Agency Contracts," *The RAND Journal of Economics*, Vol. 18, Summer 1987, pp. 296-307.

McAfee, R. P., and J. McMillan, *Incentives in Government Contracting*, University of Western Ontario Press, Ontario, Canada, 1988.

Melamud, N., and S. Reichelstein, "Value of Communication in Agencies," *Journal of Economic Theory*, Vol. 47, April 1989, pp. 334–368.

Mirrlees, J., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, Vol. 38, 1971, pp. 15–208.

Myerson, R. B., "Optimal Auction Design," *Mathematics of Operations Research*, Vol. 6, 1981, pp. 58–73.

Nalebuff, B., and J. Stiglitz, "Information, Competition, and Markets," *American Economic Review*, Vol. 73, 1983, pp. 278–283.

Osband, K., "Target Profitability Under DoD Profit Policy," unpublished RAND research, 1988a.

Osband, K., "The Role of Progress Payments in DoD Contracts," unpublished RAND research, 1988b.

Osband, K., "Contract Profitability Under DoD Profit Policy," unpublished RAND research, 1989a.

Osband, K., "Optimal Forecasting Incentives," *Journal of Political Economy*, Vol. 97, 1989b, pp. 1091–1112.

Osband, K., "Risk and Uncertainty in DoD Profit Policy," unpublished RAND research, 1990a.

Osband, K., "Quality Distortion Under Cost-Based Regulation," unpublished RAND research, 1990b.

Osband, K., and S. Reichelstein, "Information Eliciting Compensation Schemes," *Journal of Public Economics*, Vol. 27, June 1985, pp. 107–116.

Peck, M., and F. Scherer, *The Weapons Acquisition Process: Economic Analysis*, Harvard Business School, Boston, Massachusetts, 1962.

Penno, M., "Asymmetry of Pre-Decision Information and Managerial Accounting," *Journal of Accounting Research*, Vol. 22, Spring 1984, pp. 240–255.

Picard. P., "On the Design of Incentive Contracts Under Moral Hazard and Adverse Selection," *Journal of Public Economics*, Vol. 33, August 1987, pp. 305–332.

Riley, J. G., "Ex-Post Information in Auctions," *Review of Economic Studies*, Vol. 55, 1988, pp. 409–430.

Riley, J. G., and W. Samuelson, "Optimal Auctions," *American Economic Review*, Vol. 71, 1981, pp. 381–391.

Riordan, M. H., and D.E.M. Sappington, "Designing Procurement Contracts," Working Paper, 1986.

Riordan, M. H., and D.E.M. Sappington, "Awarding Monopoly Franchises," *American Economic Review*, Vol. 77, 1987, pp. 375–387.

Riordan, M. H., and D.E.M. Sappington, "Commitment in Procurement Contracting," *Scandinavian Journal of Economics*, Vol. 90, 1988, pp. 357–372.

Riordan, M. H., and D.E.M. Sappington, "Second Sourcing," *The RAND Journal of Economics*, Vol. 20, 1989, pp. 41–58.

Rogerson, W. P., "The First-Order Approach to Principal-Agent Problems," *Econometrica*, Vol. 53, 1985, pp. 1357–1368.

Rogerson, W. P., "On the Optimality of Menus of Linear Contracts," Working Paper, Northwestern University, Evanston, Illinois, 1987.

Rogerson, W. P., "Incentives, the Budgetary Process, and Inefficiently Low Production Rates in Defense Procurement," Northwestern Center for Mathematical Studies in Economics for Management Science Working Paper 851, 1989.

Rogerson, W. P., "Quality vs. Quantity in Military Procurement," *American Economic Review*, Vol. 80, 1990, pp. 83–92.

Rogerson, W. P., "Incentives, the Budgetary Process, and Inefficiently Low Production Rates in Defense Procurement" *Defence Economics*, Vol. 3, No. 1, 1991, pp. 1–18.

Rogerson, W. P., *An Economic Framework for Analyzing DoD Profit Policy*, RAND, R-3860-PA&E, 1992a.

Rogerson, W. P., *Profit Regulation of Defense Contractors and Prizes for Innovation*, RAND, R-3635-PA&E, 1992b. (A shorter version also appeared in the *Journal of Political Economy*, Vol. 97, 1989.)

Rogerson, W. P., *Overhead Allocation and Incentives for Cost Minimization in Defense Procurement*, RAND, R-4013-PA&E, 1992c.

Rogerson, W. P., "Overhead Allocation and Incentives for Cost Minimization in Defense Procurement," *The Accounting Review*, Vol. 67, October 1992d, pp. 671–690.

Rosen, S., "Contracts and the Market for Executives," Working Paper No. 3542, National Bureau of Economic Research, Cambridge, Massachusetts, 1990.

Rudin, W., *Principles of Mathematical Analysis*, 2nd edition, McGraw-Hill, New York, 1964.

Sappington, D., "Limited Liability Contracts Between Principal and Agent," *Journal of Economic Theory*, Vol. 29, 1983, pp. 1–21.

Scharfstein, D., "The Disciplinary Role of Takeovers," *Review of Economic Studies*, Vol. 55, 1988, pp. 185–199.

Scherer, F. M., *The Weapons Acquisition Process: Economic Incentives*, Harvard Business School, Boston, Massachusetts, 1964.

Schmidt, R., *Defense Profit Policy and Capital Investment*, RAND Graduate School Dissertation, 1993.

Secretary of Defense, *Annual Report to Congress FY1988*.

Smith, G. K., A. A. Barbour, T. L. McNaughter, M. D. Rich, and W. L. Stanley, *The Use of Prototypes in Weapon System Development*, RAND, R-2345-AF, March 1981.

Spence, A. M., "Job Market Signaling," *Quarterly Journal of Economics*, Vol. 87, 1973, pp. 355–374.

Stubbing, R. A., *The Defense Game*, Harper and Row, New York, 1986.

Takayama, A., *Mathematical Economics*, Cambridge University Press, Cambridge, 1985.

Tan, G., "Incentive Procurement Contracts with Costly R&D," California Institute of Technology, Social Science Working Paper 702, 1989.

Tirole, J., "Procurement and Renegotiation," *Journal of Political Economy*, Vol. 94, 1986, pp. 235–259.

Tullock, G., "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Economic Journal*, Vol. 5, 1967, pp. 224–232.

Vogelsang, I., "Regulatory Incentives Under Lumpy Capacity Investment," mimeo, Boston University, Boston, Massachusetts, 1989.

Weitzman, M. L., "The New Soviet Incentive Model," *Bell Journal of Economics*, Vol. VII, Spring 1976, pp. 251–257.

Wilson, R., "Papers on Priority Pricing," mimeo, Stanford University, Stanford California, 1988.

Wolfson, M., "Empirical Evidence of Incentive Problems and Their Mitigation in Oil and Gas Tax Shelter Programs," *Asymmetric Information, the Agency Problem and Modern Business*, Harvard Business Press, Cambridge Massachusetts, 1985.